

THESIS / THÈSE

MASTER IN BIOCHEMISTRY AND MOLECULAR AND CELLULAR BIOLOGY

Production et utilisation des microsatellites pour l'étude de la diversité intra- et inter-populations chez *Cichorium intybus* L.

MUYS, Céline

Award date:
2004

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



FACULTÉS UNIVERSITAIRES NOTRE-DAME DE LA PAIX
NAMUR

Faculté des Sciences

**PRODUCTION ET UTILISATION DES MICROSATELLITES POUR L'ETUDE DE
LA DIVERSITE INTRA- ET INTER-POPULATIONS CHEZ *Cichorium intybus* L.**

**Mémoire présenté pour l'obtention du grade de
licencié en Sciences biologiques**

Céline MUYS

Juin 2004

Production et utilisation des microsatellites pour l'étude de la diversité intra- et inter-populations chez *Chicorium intybus* L.

MUYS Céline

Résumé

La chicorée industrielle (*cichorium intybus*) est une espèce végétale dont le développement agricole est en pleine expansion en Belgique. Afin de comprendre les relations intra- et intervariétales de 6 variétés de cette espèce, des marqueurs moléculaires de type microsatellites ont été développés.

Trois loci microsatellites ont été isolés à l'aide d'un protocole d'enrichissement de banque génomique utilisant une sonde GA₁₅ biotinylée et des billes magnétiques couplées à la streptavidine.

Trois paires d'amorces correspondant aux loci microsatellites ont permis de déterminer le génotype de 54 individus appartenant à 6 variétés de chicorée industrielle : Nausica, Melci, Arancha, Madona, Vivace et 97TM17-03 (lignée obtenue par deux autofécondations successives).

Le nombre élevé d'allèles pour chaque locus (17, 10 et 14) ainsi que le pourcentage d'hétérozygotie observé (88%) supposent une diversité importante chez la chicorée industrielle. Le taux d'homozygotie dans la lignée 97TM17-03 (77,7%) est proche de la valeur attendue. Les cinq variétés (Nausica, Melci, Arancha, Madona et Vivace) sont différenciées génétiquement l'une de l'autre de manière modérée ($F_{st} = 0,09$) tandis que la lignée 97TM17-03 se détache des autres variétés de manière très importante ($F_{st} = 0,37$). Enfin, la différenciation intervariétale est influencée par la dérive génétique, par le flux de gènes mais non par les mutations ($F_{st} \neq R_{st}$). On suppose que la période séparant les variétés n'est pas suffisamment importante pour que le taux de mutation influence la différenciation génétique.

Afin de confirmer et de compléter cette étude préliminaire, l'échantillonnage ainsi que le nombre de marqueurs microsatellites doivent être augmentés.

Mémoire de licence en Sciences biologiques

Juin 2004

Promoteur: P. Van Cutsem

Remerciements

Merci à Monsieur Messiaen de m'avoir permis de réaliser ce mémoire au sein de l'unité de recherche en biologie cellulaire et moléculaire végétale.

Merci à Monsieur Van Cutsem de m'avoir donné l'opportunité de faire ce mémoire et de m'avoir encadré en tant que promoteur.

Merci tout spécialement à mon tuteur, Nicolas Dauchot. Merci pour son encadrement quotidien, son soutien et ses précieux conseils.

Merci à tous les membres du labo pour leur disponibilité et leur sympathie.

Merci à Alexia, Vincent et Yves pour tous les bons moments passés ensemble.

Merci à Olivier Hardy et Pierre Bertin d'avoir consacré une partie de leur temps pour l'explication des résultats.

Merci aux membres du jury de prendre du temps pour la lecture de mon mémoire.

Merci à toute ma famille pour leur soutien.

Merci à mes parents de m'avoir donné l'opportunité de faire ces quatre années d'étude.

Merci à ma mère pour la correction de mon mémoire.

Merci à mon frère et ma sœur d'être mon frère et ma sœur.

Merci à Vanessa pour sa présence.

Merci à Hélène pour sa motivation.

Merci à Audrey pour ses conseils.

Merci à Audrey pour ses coups de gueule.

Merci à Maxime d'avoir supporté mes nombreux sauts d'humeur. Merci d'avoir toujours été là pour me soutenir.

Enfin, merci à toutes les personnes qui, d'une façon ou d'une autre, m'ont encouragé durant ce mémoire.

Table des matières

I. INTRODUCTION GÉNÉRALE.....	4
II. INTRODUCTION	7
1. La chicorée industrielle.....	7
1.1. Introduction et généralités.....	7
1.2. Propriétés curatives de la chicorée	7
1.3. Systématique	7
1.4. Rendement industriel.....	8
2. Les marqueurs génétiques.....	8
2.1. Définition.....	8
2.1.1. Classification.....	8
2.1.2. Caractéristiques.....	8
2.2. Description et comparaison des différents marqueurs moléculaires	10
2.2.1. Les marqueurs RFLP	10
a. Principe	10
b. Technique.....	10
c. Avantages et inconvénients.....	12
2.2.2. Les marqueurs de type PCR.....	12
a. Les microsatellites ou SSR.....	12
a.1. Principe	12
a.2. Technique.....	12
a.3. Avantages et inconvénients.....	14
b. La technique AFLP	14
b.1. Principe	14
b.2. Technique.....	14
b.3. Avantages et inconvénients.....	14
c. La technique RAPD	16
c.1. Principe	16
c.2. Avantages et inconvénients.....	16
2.2.3. Les marqueurs SNP.....	16
a. Principe	16
b. Technique.....	18
b.1. Mise en évidence des SNP dans le génome.....	18
b.2. Génotypage à l'aide des SNP.....	18
2.3. Application des marqueurs moléculaires.....	21
2.3.1. Empreinte génétique	21
a. Principe	21
b. Applications	21
2.3.2. Cartes génétiques	21
a. Cartographie des marqueurs moléculaires	21
b. Cartographie d'un gène majeur.....	23
c. Cartographie d'un caractère quantitatif	23
3. Les microsatellites	24
3.1. Définition.....	24

3.2. Différence de distribution des microsatellites entre les régions codantes et non codantes.....	25
3.3. Conservation des marqueurs microsatellites entre les espèces.....	27
3.4. Modèles et taux de mutation des microsatellites.....	27
3.4.1. Modèles.....	27
3.4.2. Taux de mutation.....	28
3.5. Cycle de vie d'un microsatellite.....	28
3.5.1. Naissance.....	29
3.5.2. Croissance.....	27
3.5.3. Mort et résurrection.....	27
3.6. Exemples d'applications des marqueurs microsatellites.....	30
3.6.1. Création de groupe de liaison.....	30
3.6.2. Utilisation de microsatellites pour la traçabilité de la filière animale.....	30
3.6.3. Utilisation de microsatellites pour la détermination de l'introgression au sein d'une espèce végétale.....	30
3.6.4. Exemple d'utilisation des microsatellites pour l'identification de DTL.....	31
3.6.5. Les microsatellites en génétique des populations.....	31
4. Stratégie pour l'isolation des marqueurs moléculaires.....	35
4.1. Méthode originale.....	35
4.2. Méthode utilisant l'extension d'amorces.....	35
4.2.1. Méthode de Ostender.....	37
4.2.2. Méthode de Paetkau.....	37
4.3. Isolation par hybridation sélective.....	37
5. Objectifs.....	39

III. MATERIEL ET METHODE.....42

1. Extraction d'ADN génomique de chicorée.....	42
2. Génération d'une collection de fragments d'ADN génomique.....	43
2.1. Restriction de l'ADN génomique de chicorée.....	43
2.2. Traitement à la nucléase.....	43
2.3. Traitement à la phosphatase alcaline.....	43
2.4. Récupération des fragments d'ADN entre 400 et 900 pb avec le kit QiaExII.....	44
2.5. Hybridation des amorces SNX pour former l'adaptateur double brin.....	44
2.6. Réaction de ligation/restriction simultanée de l'adaptateur SNX aux fragments génomique.....	44
3. Enrichissement des fragments génomiques en séquences contenant des motifs microsatellites.....	45
3.1. Hybridation soustractive.....	45
3.2. Insertion des fragments d'ADN gen. enrichis en microsatellites dans le Vecteur pCR®II et transformation dans TOP10F'.....	46
3.3. Criblage sur membrane de nylon.....	47
4. Séquençage des clones positifs.....	49
4.1. Miniprep.....	49
4.2. Séquençage.....	50
5. Design des amorces.....	51

IV. RESULTATS ET DISCUSSION.....	54
1. Clonage de fragments d'ADN génomique enrichis en microsatellites.....	54
1.1. Isolement de l'ADN génomique.....	54
1.2. Traitement de l'ADN génomique en vue de lier les adaptateurs SNX aux fragments d'ADN.....	54
1.3. Enrichissement de la collection de fragments d'ADN en séquences microsatellites	56
2. Détermination des amorces.....	60
2.1. Sélection des clones à séquencer.....	60
2.2. Séquençage des clones sélectionnés.....	61
2.2. Séquençage des clones sélectionnés.....	62
2.3. Design des amorces	64
2.4. Test d'amplification.....	66
2.5. Etude du polymorphisme.....	67
3. Analyse du polymorphisme intravariétal et intervariétal.....	69
3.1. Extraction d'ADN	69
3.2. Détermination de la taille des allèles.....	69
3. Interprétation des résultats.....	73
4. Discussion générale	73
V. CONCLUSIONS ET PERSPECTIVES.....	89
VI. BIBLIOGRAPHIE	91

Abréviations

- ✓ #: référence
- ✓ %: pourcent
- ✓ °C: degré Celsius.
- ✓ µg: microgramme
- ✓ µl: microlitre
- ✓ µM: micromolaire
- ✓ A: Adenine
- ✓ ADN_{gén.}: Acide désoxyribonucléique génomique
- ✓ ADN: Acide Désoxyribonucléique
- ✓ ADNc: ADN complémentaire
- ✓ AFLP : Amplified Fragment Length Polymorphism
- ✓ ARN_m: Acide Ribonucléique messagé
- ✓ C: Cytosine
- ✓ cm: centimètre
- ✓ cM: centimorgan
- ✓ Fig.: Figure
- ✓ Fw: Forward
- ✓ G: Guanine
- ✓ gr: gramme
- ✓ h: heure
- ✓ Ha: hectare
- ✓ Ind.: individu
- ✓ Kb : kilo bases
- ✓ LB: Luria Broth
- ✓ M: molaire
- ✓ mA: mili ampere.
- ✓ mg: milligramme
- ✓ ml: millilitre
- ✓ mM: millimolaire
- ✓ MS: Murashige Skoog
- ✓ NCBI: National Center for Biotechnology Information
- ✓ ng : nanogramme
- ✓ nt: nucleotide.
- ✓ OGM: Organisme Génétiquement Modifié
- ✓ p/v: poid/volume
- ✓ pb: paire de base
- ✓ PCR : Polymerase Chain Reaction
- ✓ PW: poids moléculaire
- ✓ QTL : Quantitative Trait Loci
- ✓ RAPD : Randomly Amplified Polymorphic DNA
- ✓ RFLP : Restriction Fragment Length Polymorphism
- ✓ Rpm: tour par minute
- ✓ RRS : Reduced Representation Shotgun
- ✓ Rv: Reverse
- ✓ SBE: Single Base Extension
- ✓ SL: Smart Ladder
- ✓ SNP : Single Nucléotide Polymorphism
- ✓ SSR : Simple Sequence Repeat
- ✓ T: Tyrosine
- ✓ Tm: melting temperature
- ✓ U: Unité
- ✓ UV: ultraviolet
- ✓ v/v: volume/volume
- ✓ VNTR : Variable Number of Tandem Repeat

I. Introduction générale

La chicorée industrielle est une plante bisannuelle appartenant à la famille des Astéracées. On y distingue trois sous-espèces: *Cichorium intybus sylvestre*, la chicorée sauvage, *Cichorium intybus foliosum* plus communément appelée «chicon» et *Cichorium intybus sativum*, la chicorée industrielle.

La chicorée industrielle se distingue des autres sous-espèces par la teneur élevée de sa racine en inuline, un polymère pouvant contenir 50 unités fructosyles et terminé par un glucose : $(G1)-2F(1-2F)_n$.

Du fait de sa non métabolisation par l'organisme, l'inuline est à l'heure actuelle de plus en plus utilisée dans le secteur agro-alimentaire en tant que fibre, agent de texture, substitut de graisses ou édulcorant basse calories pour les plus petits fragments.

Récemment, de nombreuses applications non-agroalimentaires ont été proposées: on a en effet observé que l'inuline régule le transit intestinal, favorise l'excrétion de phénols neutres et influence positivement différents paramètres sanguins (Guldentops, 1994). L'inuline favorise également le développement de bifidobactéries bénéfiques, qui, par leur activité, produisent des métabolites secondaires réduisant les risques de cancer du colon.

Face à une demande croissante, la culture de la chicorée industrielle est en pleine expansion dans notre pays. Un recensement agricole réalisé en 2003 par l'Institut National des Statistiques (Belgique) montre que la surface agricole occupée par la chicorée a augmenté de 18,4% par rapport à 2002 (13.346 ha en 2002 et 15.797 ha en 2003).

La sélection de la chicorée à inuline est relativement récente puisque celle-ci débuta dans le courant des années 80. Auparavant, la sélection était en effet principalement orientée vers la production de "chicorée à café" et de chicorée witloof.

Jusqu'à présent, les sélections ont classiquement été menées sur des bases strictement phénotypiques. Cependant, grâce à l'apparition des marqueurs moléculaires, de nouvelles méthodes de sélection sont entrain de voir le jour.

Pour rappel, les marqueurs moléculaires sont des loci polymorphes permettant de donner une information sur le génotype de l'individu étudié (B. Najimi *et al.*, 2003). Ainsi, le développement de QTL, d'empreintes génétiques et les rétrocroisements assistés par marqueurs permettent de sélectionner les individus de manière plus rationnelle.

D'un point de vue fondamental, les marqueurs moléculaires rendent possible la création de groupes de liaison, la cartographie de gènes ainsi que l'étude de la distribution génotypique d'une population afin de mieux comprendre son organisation et ses caractéristiques (C. Dagorn, 1999; D. de Vienne, 1998).

L'objectif de ce mémoire est de réaliser une étude de la diversité génétique intra- et inter-variétale au sein de six cultivars de chicorée industrielle (*Cichorium intybus*). Afin de réaliser cette étude, il a fallu faire le choix d'un marqueur. Notre choix s'est porté vers les microsatellites.

Avec l'AFLP (Amplified Fragment Length polymorphism) et les SNP (Single Nucleotide Polymorphism), les microsatellites font partie des marqueurs les plus utilisés actuellement.

Une séquence microsatellite est un segment répété de 2 à 6 nucléotides formant un fragment plus ou moins uniforme de 100 nt et présent principalement dans les régions non codantes du génome (G. K. Chambers *et al.*, 2000).

Le principe de ces marqueurs se base sur le fait que, d'un individu à l'autre, le nombre de répétitions microsatellites pour un locus donné peut varier. Ainsi, chaque variation du nombre de répétitions pour ce locus est défini comme étant un allèle distinct. Afin de mettre en évidence ces différents allèles, on réalise une amplification à l'aide d'amorces s'hybridant dans les régions bordant le motif microsatellite (Morgan & Olivieri, 1993).

Les marqueurs microsatellites possèdent plusieurs qualités. Contrairement à l'AFLP, cette technique est codominante, c'est-à-dire que les deux allèles d'un individu diploïde peuvent être identifiés.

A l'inverse des SNP, aucune information préalable relative à la séquence du génome de la plante étudiée (séquençage systématique ou EST) n'est nécessaire.

De plus, le polymorphisme des microsatellites est généralement important, vu la valeur élevée du taux de mutation qui les affecte. La localisation prédominante de ces séquences dans les régions non codantes du génome implique que ces marqueurs sont majoritairement neutres. Enfin, une fois les mises au point réalisées, leur utilisation est simple, robuste et les résultats sont reproductibles.

Au cours de ce mémoire, nous tenterons d'isoler un maximum de séquences microsatellites polymorphes et nous utiliserons les marqueurs mis au point pour analyser la diversité intra- et inter-génétique de cinq variétés cultivées de chicorée industrielle (Nausica, Arancha, Melci, Madona, Vivace) ainsi que d'une lignée autofécondée en I2 (97TM17-03).

II. Introduction

1. La chicorée industrielle

1.1. Introduction et généralités

La chicorée industrielle (*Cichorium intybus*) (Fig.1, Fig. 2) est une espèce végétale bisannuelle, pouvant atteindre 80 à 90 cm de haut. C'est une plante diploïde ($2n=18$) et allogame.

La chicorée est connue de l'homme depuis des milliers d'années. La première utilisation de cette plante fut pour la production de papyrus par les Egyptiens. Plus tard, la chicorée fut renommée pour ses propriétés dépuratives et rafraîchissantes ainsi que pour son effet apaisant pour l'estomac, le foie et le cœur. Ce n'est qu'au 17^e siècle que l'on découvrit ses propriétés nutritives. Au 18^e siècle, le blocus continental imposé par Napoléon I^{er} afin d'affaiblir l'Angleterre fit grimper le prix du café de façon exorbitante. La chicorée s'est alors avérée être la plante qui le remplaçait au mieux. Au 19^e siècle, l'industrie liée à la culture de la chicorée connut son heure de gloire. Cependant, jusqu'en 1980, la chicorée fut uniquement utilisée pour la production de substitut de café. Au-delà de cette date, elle a également été utilisée pour la production d'inuline, chaîne d'oligo-fructane, fortement utilisée dans le secteur agro-alimentaire en tant que substitut du saccharose (P. Cornelis, 2003).



Fig. 1. *Cichorium intybus*
Inflorescence

1.2. Propriétés curatives de la chicorée

La racine de chicorée est actuellement utilisée pour augmenter le taux de bactéries bénéfiques dans l'intestin, abaisser le taux de lipide sanguins, stimuler l'appétit, soulager les troubles gastriques fonctionnels et contre la constipation.

1.3. Systématique

La systématique de la chicorée peut être définie comme suit

Embranchement :	Spermatophytes
Sous-embranchement :	Angiospermes
Classe :	Dicotylédones
Ordre :	Astérales
Famille :	Astéracées
Sous-famille :	Lactucoidea ou Cichorioideae
Genre :	<i>Cichorium</i>
Espèce :	<i>intybus</i> Linneus

Dans l'espèce *intybus* L., on retrouve trois variétés :

- *Cichorium intybus sylvestre*, la chicorée sauvage,
- *Cichorium intybus foliosum* plus communément appelée «chicon »,
- *Cichorium intybus sativum*, la chicorée industrielle, objet de cette recherche.



Fig. 2. *Cichorium intybus*
Parties aériennes à l'étape
végétative

1.4. Rendement industriel

La chicorée se cultive dans de nombreux pays dont notamment le Royaume-Uni, la Belgique, la France, les Pays-Bas, l'Allemagne, l'Afrique de sud, les Etats-Unis, l'Australie et l'Inde. Un recensement agricole réalisé en 2003 en Belgique par l'Institut National des Statistiques montre que la surface agricole occupée par la chicorée a augmenté de 18,4% par rapport à 2002 (13346 ha en 2002 et 15797 ha en 2003). La culture de cette plante est donc en pleine expansion dans notre pays.

2. Les marqueurs génétiques

2.1. Définition

Un marqueur génétique est un locus polymorphe qui renseigne sur le génotype de l'individu qui le porte (B. Najimi *et al.*, 2003).

2.1.1. Classification

On peut distinguer plusieurs types de marqueurs génétiques, principalement regroupés en deux classes (C. Dagorn, 1999) :

- **Les marqueurs biochimiques**

Ces marqueurs correspondent à des protéines quasi identiques mais présentant un polymorphisme ponctuel dans la séquence en acides aminés chez différents individus. Les marqueurs biochimiques les plus connus sont les isozymes. Ils correspondent aux diverses formes d'une enzyme, possédant la même fonction mais différant par la structure polypeptidique (C. François, 1994). Ils permettent de déterminer la présence de l'allèle correspondant à chacune des ces variations. Ils reflètent donc le polymorphisme des séquences codantes du génome.

- **Les marqueurs moléculaires d'ADN**

Ils mettent en évidence la présence de polymorphisme chez différents individus au niveau de séquences génomiques codantes ou non. Les plus connus sont les marqueurs RFLP, RAPD, AFLP, microsatellites et SNP.

2.1.2. Caractéristiques

Un marqueur génétique devrait idéalement posséder les propriétés suivantes. Aucun type de marqueur ne les possède toutes mais il est évident que plus on regroupe de caractéristiques, plus les conditions d'étude sont optimales (D. de Vienne, 1998)

- **Neutralité** : les différents allèles ne doivent pas avoir d'effet sur le phénotype de l'individu. Dans le cas des marqueurs biochimiques, les substitutions alléliques au locus marqueur n'ont pas d'autre effet phénotypique (et donc éventuellement sélectif) que ceux qui permettent de déterminer son génotype.
- **Polymorphisme** : un marqueur possédant de nombreux allèles permet de mieux mettre en évidence la variabilité entre individus.
- **Codominance** : l'hétérozygote présente simultanément les caractères de ses deux parents. On peut donc distinguer l'homozygote qui ne présente qu'un seul caractère,

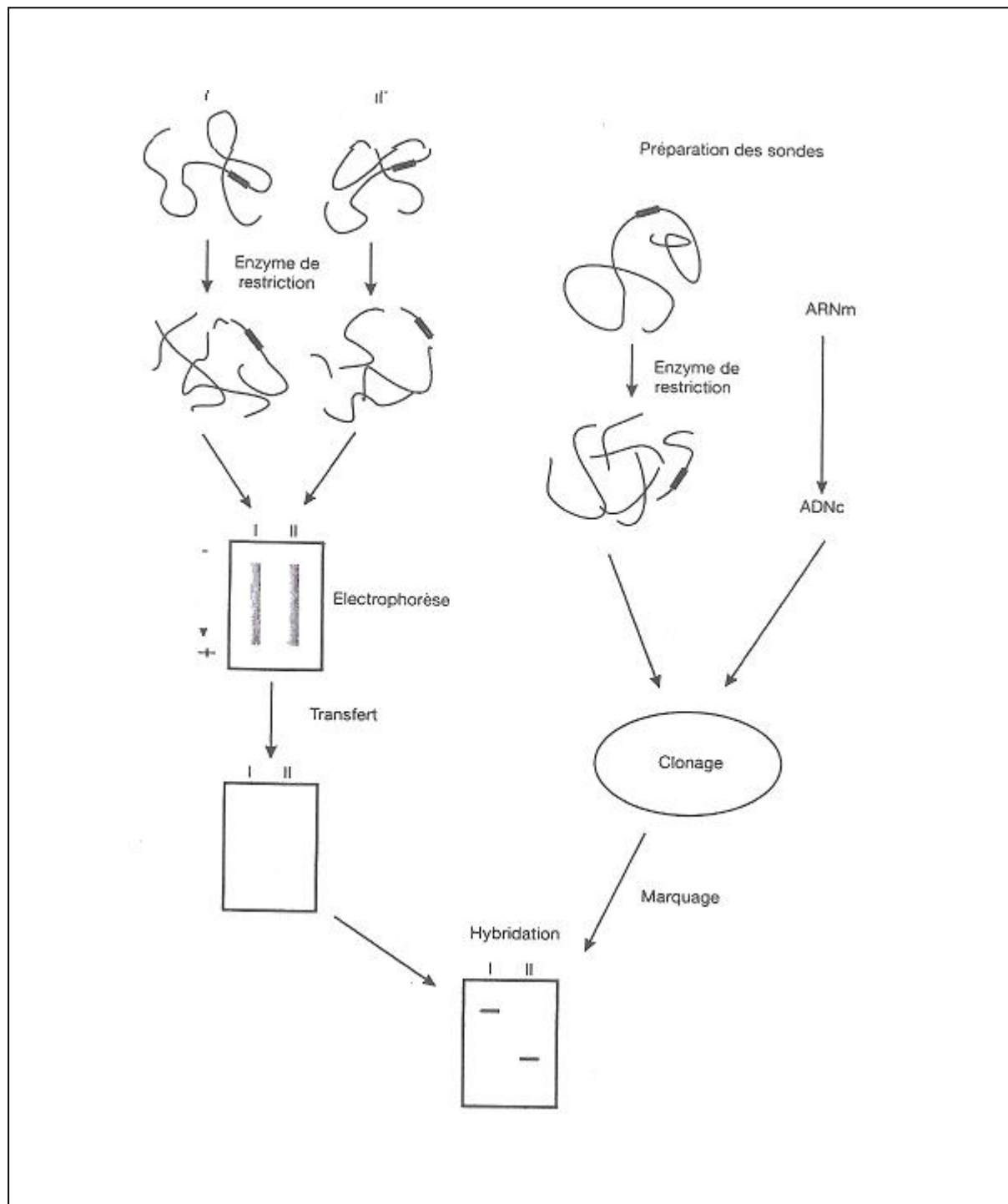


Fig. 3. (D. de Vienne, 1998) Principe de la révélation du polymorphisme de longueurs des fragments de restriction (RFLP). I et II représentent deux individus. La molécule d'ADN est schématisée par un filament. Le rectangle gris sur l'ADN représente une région particulière reconnue par la sonde. Après digestion, le fragment portant cette région est plus long chez l'individu I que chez l'individu II. Cette différence de longueur sera révélée par hybridation de la sonde marquée sur l'ADN après migration. La sonde est obtenue soit à partir d'ADN génomique soit à partir d'ARNm.

de l'hétérozygote qui en présente deux (dans le cas d'un marqueur dominant, seul un caractère peut être visualisé quelles que soient les conditions).

- Non-épistasie : dans le cas des marqueurs biochimiques, le génotype peut être lu à partir de son phénotype quel que soit le génotype des autres locus. Il n'y a pas d'interaction entre les locus.
- Insensibilité aux variations spatiales et temporelles : le marqueur doit être observable dans n'importe quel milieu, tout au long d'une expérience, à n'importe quel stade de développement et dans tous les organes.

Une brève description des différents marqueurs moléculaires sera développée dans ce travail ainsi que leurs applications en biologie moléculaire

2.2. Description et comparaison des différents marqueurs moléculaires

Contrairement aux marqueurs morphologiques et biochimiques, les marqueurs moléculaires ne sont pas influencés par les fluctuations de l'environnement et sont indépendants de l'organe utilisé et du stade de développement de la plante (sauf si le travail est effectué à partir d'ADNc)

Les méthodes de marquage moléculaire peuvent être regroupées en deux grandes catégories (B. Najimi *et al.*, 2003) :

- les marqueurs de type RFLP (Restriction Fragment Length Polymorphism) ;
- les marqueurs basés sur la méthode PCR.

2.2.1. Les marqueurs RFLP

a. Principe

La technique RFLP, développée par Botstein *et al.* (1980), repose sur la mise en évidence du polymorphisme de longueur de fragments obtenus suite à l'action d'enzymes de restriction spécifiques.

Le principe est le suivant : les enzymes de restriction coupent l'ADN en des sites spécifiques appelés sites de restriction. La spécificité de ces sites est telle que le remplacement d'une seule base suffit à empêcher la coupure de l'ADN par l'enzyme utilisée. Donc, une présence-absence de site de restriction entraîne un polymorphisme de longueur de fragment. Ce phénomène n'étant pas rare, la digestion de l'ADN de deux individus quelconques au sein d'une même espèce produit de très nombreuses différences de longueur des fragments.

b. Technique

Pour détecter ce polymorphisme, une multitude de fragments d'ADN est générée par digestion enzymatique. Ceux-ci sont ensuite séparés sur gel d'agarose et transférés par capillarité sous forme dénaturée sur une membrane de nylon. Le polymorphisme est révélé par Southern blot. Les fragments homologues seront mis en évidence grâce à un fragment d'ADN sonde déterminé. (Fig. 3)

C'est le couple enzyme/sonde qui constitue le marqueur.

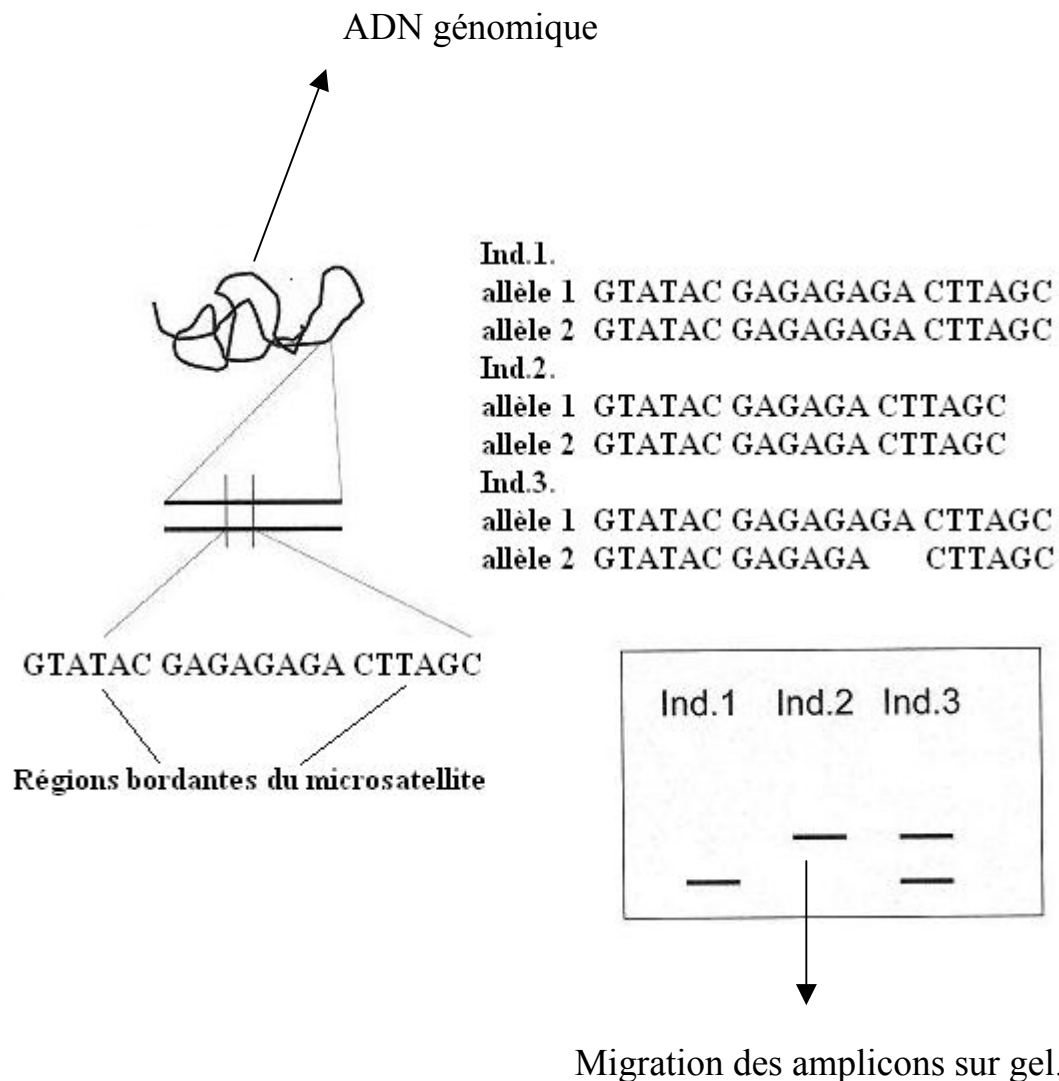


Fig. 4. Principe des marqueurs microsatellites. Les régions bordantes d'un locus microsatellite sont identifiées. Celles-ci sont ensuite amplifiées chez les individus étudiés (Ind.1, Ind. 2, Ind. 3). Le polymorphisme de taille est révélé par migration des amplicons. Pour un individu diploïde hétérozygote (Ind. 3), les deux allèles sont visibles.

c. Avantages et inconvénients

Bien que cette technique soit codominante et permette une analyse génétique complète, elle est lente et laborieuse. Les étapes de transfert et d'hybridation empêchent une automatisation du travail (D. de Vienne, 1998, B. Najimi *et al.*, 2003).

2.2.2. Les marqueurs de type PCR

Le développement de la technique PCR offre l'avantage d'analyser des marqueurs moléculaires en un temps court tout en permettant l'utilisation de faibles concentrations d'ADN. Les marqueurs de type PCR tendent à remplacer les marqueurs morphologiques, biochimiques et RFLP. Les plus largement utilisés sont les microsatellites ou SSR (Simple Sequence Repeat), AFLP (Amplified Fragment Length Polymorphism) et la RAPD (Randomly Amplified Polymorphic DNA) (D. de Vienne, 1998, B. Najimi *et al.*, 2003).

a. Les microsatellites ou SSR

Cette technique a été décrite par Morgan et Olivieri en 1993.

a.1. Principe

Un microsatellite est un ensemble de segments répétés de 2 à 6 nt formant un fragment plus ou moins uniforme d'environ 100 nt de long et présent principalement dans les régions non codantes de l'ADN (G. K. Chambers *et al.*, 2000). On parle aussi de séquences répétées en tandem ou SSR (Single Sequence Repeat). Chez les végétaux supérieurs, on estime qu'il y aurait en moyenne un microsatellite tous les 30 à 100 kb (D. de Vienne, 1998).

D'un individu à l'autre, le nombre de répétitions dans un microsatellite peut varier. Il est donc possible de distinguer deux individus sur base d'une variation du nombre de répétitions qu'ils possèdent pour un microsatellite donné.

Les régions bordant un locus microsatellite donné doivent être identiques chez tous les individus à tester. On utilisera ces séquences afin de définir les amorces nécessaires à l'amplification d'un microsatellite particulier. C'est donc la paire d'amorces spécifiques des bordures 5' et 3' du microsatellite qui constitue le marqueur (Fig. 4).

a.2. Technique

Les fragments d'amplification sont révélés par différentes méthodes :

- **L'électrophorèse.** On peut observer le polymorphisme de taille entre différents individus par comparaison entre la distance de migration des amplicons dans un gel d'agarose ou d'acrylamide.
- **La spectrométrie de masse.** Deux amplicons de masses différentes suggèrent que ceux-ci sont de tailles différentes.
- **Le séquençage.** Permet une analyse directe du polymorphisme par comparaison des séquences.

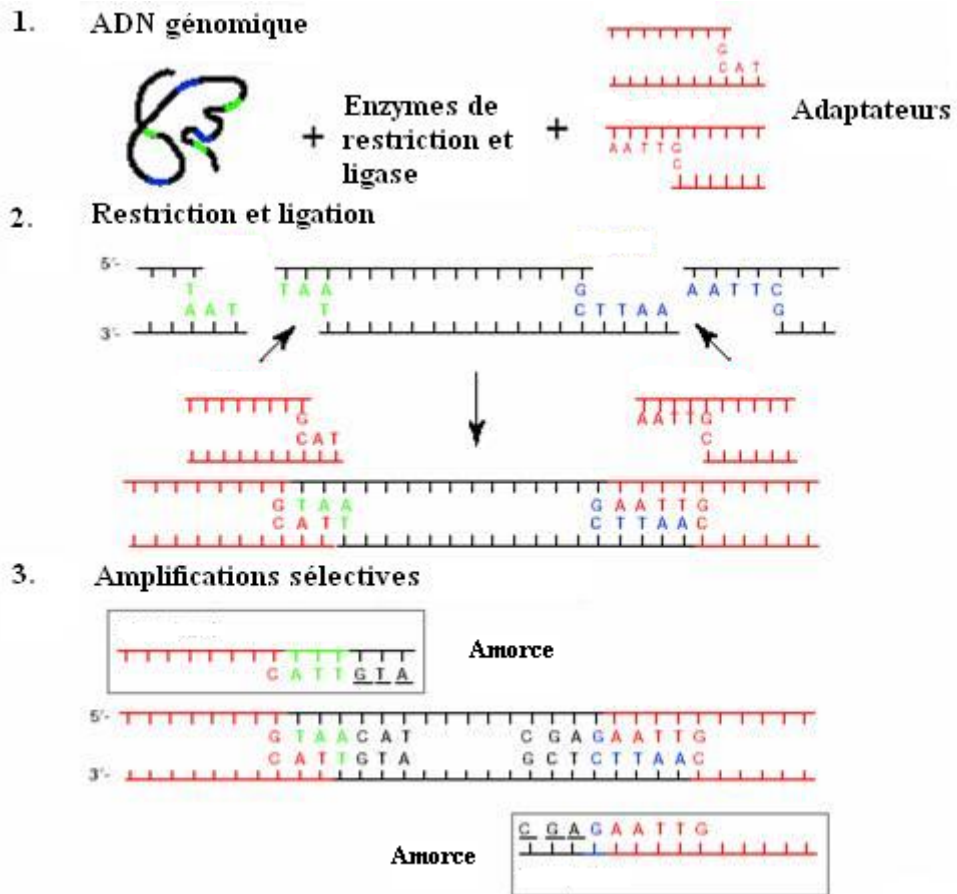


Fig. 5. Principe des marqueurs AFLP. 1. L'ADN génomique est restreint à l'aide de deux enzymes de restriction. 2. Les adaptateurs se lient aux fragments d'ADN. 3. Les bases arbitraires permettent une amplification sélective de certains fragments.

a.3. Avantages et inconvénients

L'inconvénient majeur de cette technique est la préparation assez lourde des marqueurs microsatellites. Il faut en effet :

- cribler une banque génomique, enrichie ou non, avec une sonde microsatellite,
- séquencer les clones positifs,
- définir et synthétiser les amorces oligonucléotidiques,
- tester les paires d'amorces dans un échantillon d'individus afin d'évaluer la fonctionnalité des amorces définies et le polymorphisme des fragments amplifiés.

En revanche, les marqueurs sélectionnés ont comme avantages d'être spécifiques de locus, codominants et polymorphes. De plus, leur utilisation repose sur une simple PCR.

Ces marqueurs seront revus en détail dans le point 3.

b. La technique AFLP

b.1. Principe

Développée par Vos *et al.* en 1995, cette technique est fondée sur la mise en évidence conjointe de polymorphisme de sites de restriction et d'hybridation d'amorces arbitraires. Cette technique utilise à la fois les enzymes de restriction et l'amplification PCR.

b.2. Technique

La mise en évidence du polymorphisme se fait en plusieurs étapes (Fig. 5) :

1. L'ADN génomique est clivé par deux enzymes de restriction.
2. Des adaptateurs de séquences connues sont ajoutés aux extrémités des fragments de restriction générant ainsi une matrice pour l'amplification.
3. Une pré-amplification est réalisée à l'aide d'amorces de séquences complémentaires à la séquence des adaptateurs et des sites de restriction. Cette amorce est prolongée en 3' par une base arbitraire.
4. La deuxième amplification, dite sélective, utilise des amorces identiques aux premières mais prolongées à l'extrémité 3' par deux ou trois nucléotides arbitraires. Ces amorces sélectives permettent de réduire le nombre de fragments amplifiés à une centaine.
5. Ces fragments sont séparés par électrophorèse sur gel d'acrylamide dénaturant puis visualisés par coloration au nitrate d'argent ou révélés grâce à un marquage radioactif ou fluorescent réalisé lors de l'amplification sélective.

C'est la combinaison enzyme de restriction/amorce qui permet de révéler le polymorphisme entre les individus et constitue donc le marqueur AFLP.

b.3. Avantages et inconvénients

Cette technique est puissante, reproductible et rapide. En outre, l'AFLP ne nécessite aucune connaissance préalable de séquence du génome de la plante étudiée ni de construction de banques génomiques ou cDNA, contrairement aux SSR et RFLP (en effet, une banque génomique est souvent nécessaire pour la construction d'amorces dans le cas des microsatellites et pour la production de sondes dans le cas des marqueurs RFLP). Toutefois, ces marqueurs sont dominants.

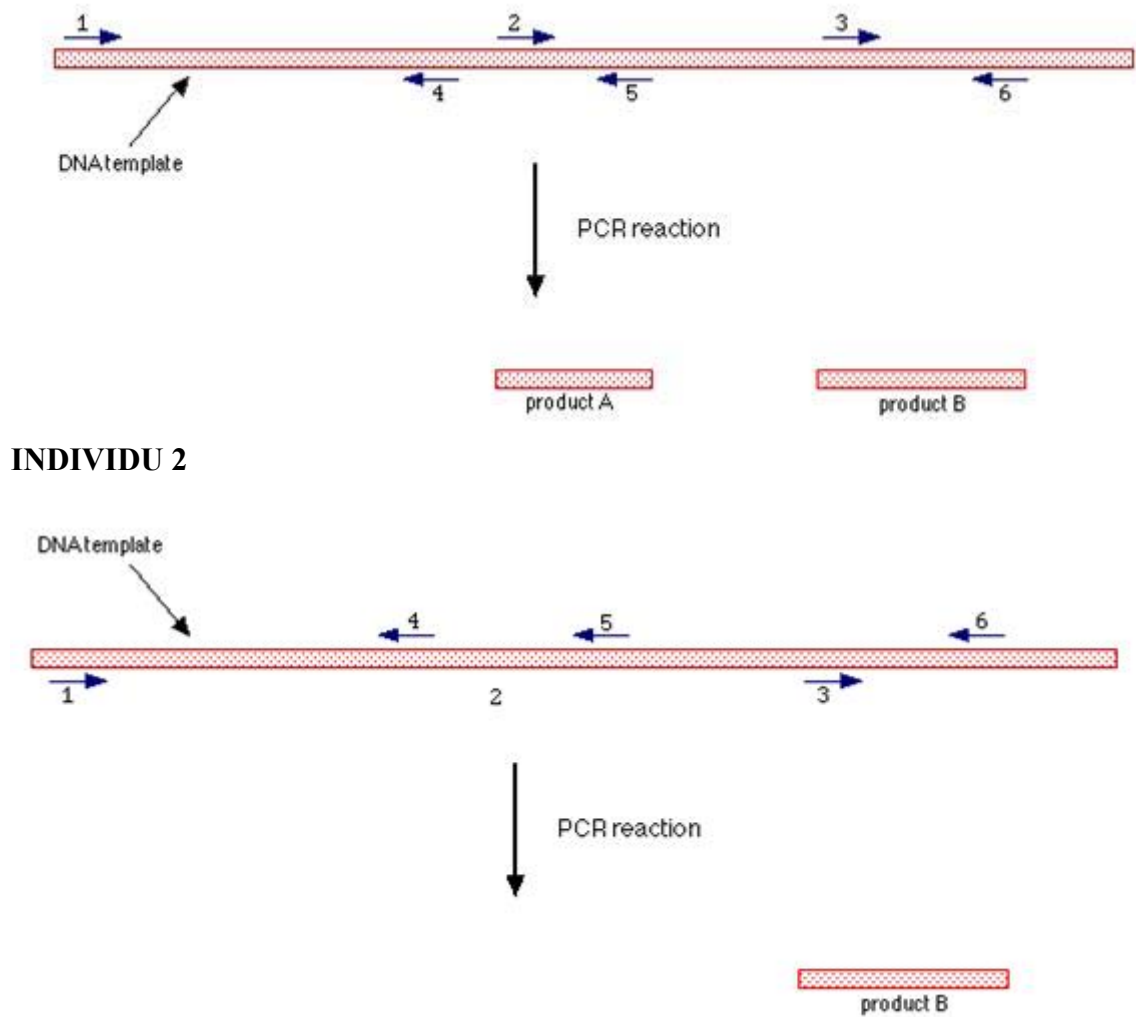


Fig. 6. Principe des marqueurs RAPD. Les amorces arbitraires s'hybrident sur l'ADN. L'individu 1 aura deux amplicons. Chez l'individu 2, le site d'hybridation 2 possède une mutation. Seul un amplicon sera détecté.

c. La technique RAPD

Cette technique a été décrite par William *et al.* en 1990.

c.1. Principe

Cette technique consiste en l'amplification par PCR de fragments d'ADN génomique en utilisant des amorces arbitraires de courte taille (10 pb) allant s'hybrider au hasard dans le génome. Si deux sites d'hybridation sont suffisamment proches et situés sur chacun des deux brins d'ADN, il y aura amplification. En revanche si deux sites de fixation d'amorce sont trop éloignés, l'amplification n'aura pas lieu.

Le polymorphisme décelé est dû à des mutations au niveau des sites de fixation des amorces qui ne reconnaissent plus leur site d'hybridation et ne peuvent plus s'y fixer. On observera donc une bande chez un individu non muté et l'individu muté n'aura pas de bande correspondante dans le gel (Fig. 6). Les produits d'amplification sont généralement visualisés par électrophorèse sur gel d'agarose. Les amorces constituent les marqueurs.

c.2. Avantages et inconvénients

Cette technique est simple, rapide et ne nécessite ni restriction, ni transfert sur membrane, ni marquage radioactif, ni connaissance préalable de la séquence nucléotidique. Néanmoins, la RAPD manque de reproductibilité vu sa sensibilité à la concentration en ADN et aux conditions d'amplification. Cette technique est aujourd'hui virtuellement abandonnée.

2.2.3. Les marqueurs SNP

a. Principe

Un nouveau type de marqueur a pu être développé grâce à l'essor de la bioinformatique et du séquençage systématique des génomes : les SNP (Single Nucleotide Polymorphism). Ces marqueurs sont des mutations ponctuelles d'une seule base au sein d'une séquence chez différents individus. Les SNP sont majoritairement bialléliques car dans la plupart des cas, seuls deux nucléotides peuvent être présents à une position donnée. Deux explications ont été proposées pour comprendre ce phénomène.

La première tient compte du fait que la fréquence de substitution de nucléotide unique à l'origine des SNP est très faible. On estime cette fréquence entre 1.10^{-9} et 5.10^{-9} par nucléotide et par an pour une position neutre chez les mammifères. De ce fait, la probabilité que 2 changements de base indépendants se produisent à une même position est extrêmement faible.

La deuxième explication se base sur le biais des mutations. Les mécanismes de mutation peuvent être de deux types :

- Soit une transition c'est-à-dire un échange purine-purine ($A \rightarrow G$) ou un échange pyrimidine-pyrimidine ($C \rightarrow T$).
- Soit une transversion c'est-à-dire un échange purine-pyrimidine ou pyrimidine-purine.

Il a été montré que le biais des mutations tourne en faveur de la transition. Les deux facteurs expliqueraient que les SNP soient bialléliques.

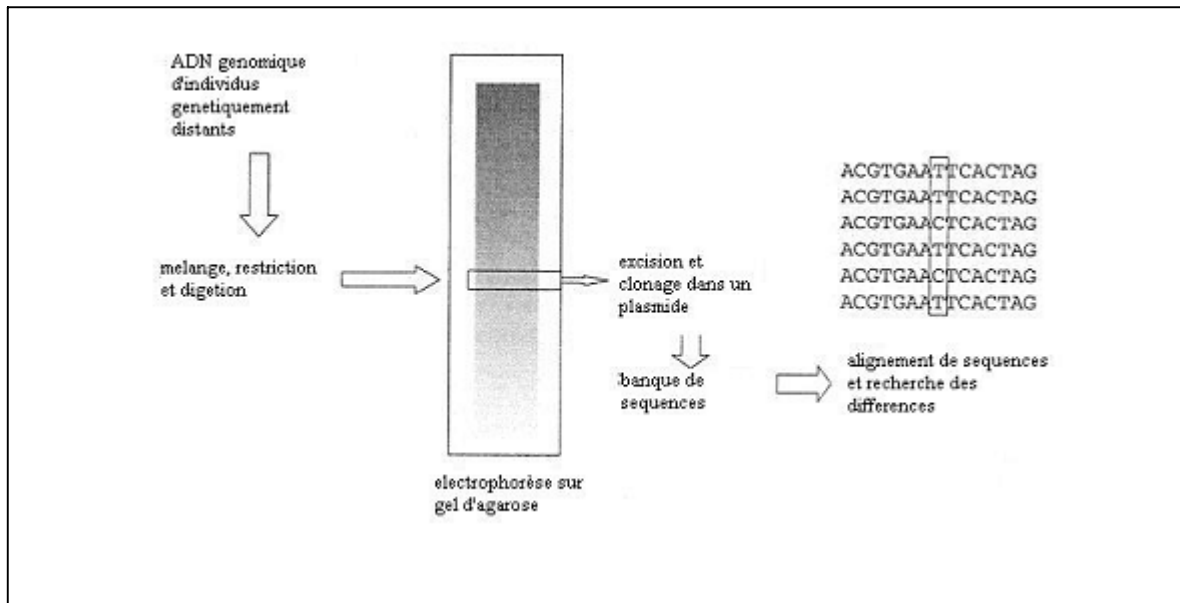


Fig. 7. Principe du Reduced Representation Shotgun (RRS). Cette technique permet d'isoler les SNP au sein d'un génome (A. Vignal *et al.*2002)

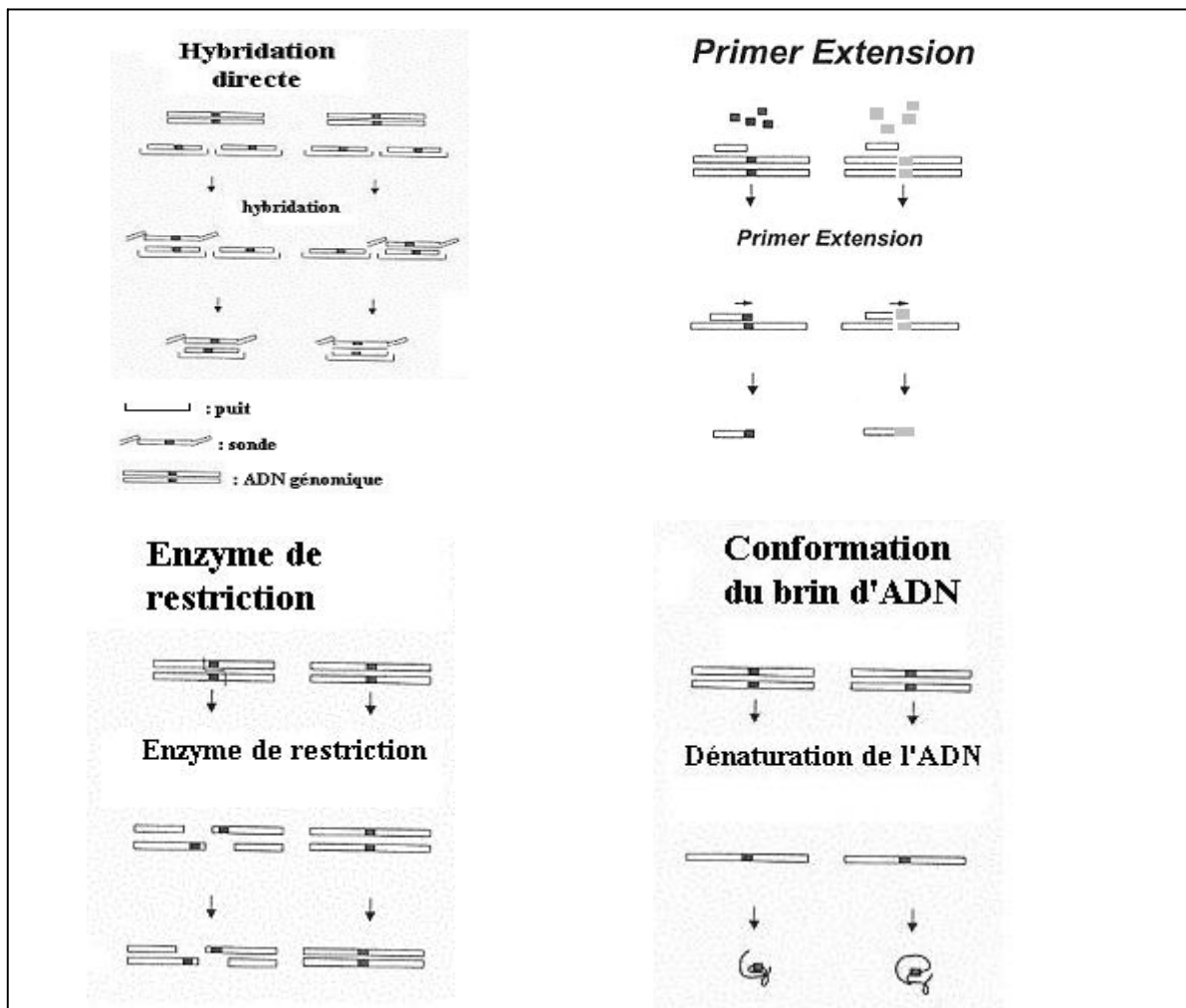


Fig. 8. Techniques utilisées pour le génotypage à l'aide des SNP. (A. Vignal *et al.*2002)

b. Technique

b.1. Mise en évidence des SNP dans le génome.

Pour utiliser les SNP comme marqueurs moléculaires il faut tout d'abord les isoler dans le génome. Les deux techniques les plus utilisées sont la comparaison de séquences spécifiques de locus et le RRS c'est-à-dire Reduced Representation Shotgun.

La comparaison de séquences spécifiques de locus est en fait un séquençage direct de produits PCR obtenus chez différents individus. Il est donc nécessaire de déterminer au préalable des amorces spécifiques de locus. La difficulté de cette méthode intervient lors de l'interprétation du séquençage. En effet, il faut distinguer les artéfacts de séquençage et le polymorphisme.

La méthode RRS se déroule en plusieurs étapes. Tout d'abord, l'ADN de différents individus est mélangé. Une restriction et une migration sur gel d'agarose vont être effectuées. Ensuite, une représentation réduite de ces génomes est réalisée en sélectionnant des fragments d'une certaine taille. Ces fragments sont insérés dans des plasmides et forment donc une banque de séquence réduite. La dernière étape est un alignement de séquences et un repérage des mutations uniques dans ces séquences (Fig. 7.).

b.2. Génotypage à l'aide des SNP

Plusieurs techniques sont utilisées afin de mettre le polymorphisme de ces marqueurs en évidence (Fig. 8.) :

- **Hybridation directe** : cette technique consiste à repérer les SNP à l'aide d'une sonde.
- **Restriction enzymatique** : seuls les SNP se trouvant dans les sites de restriction peuvent être détectés. La mutation empêchera l'enzyme de couper dans le site et on observera des fragments de longueurs différentes pour deux individus polymorphes.
- **Analyse de la conformation de l'ADN simple brin** en condition non dénaturante. Un changement d'une seule base peut modifier la conformation de l'ADN simple brin.
- **Extension d'amorce**. Cette technique se base sur la PCR. L'analyse des SNP peut être réalisée de deux manières :
 - Deux amorces, chacune spécifique d'un des deux allèles, sont associées à un fluorochrome de couleur différente. Après PCR, la fluorescence émise permettra de connaître l'allèle amplifié.
 - Une amorce SBE (Single Base Extension) bordant le SNP, n'ajoute qu'un seul nucléotide à l'extrémité 3'. Chaque nucléotide est associé à un fluorochrome différent. Il est donc aisé, par la détection de la fluorescence après amplification, de connaître l'allèle de l'individu étudié.

L'extension d'amorces est à l'heure actuelle la technique la plus répandue.

(A. Vignal *et al.*, 2002)

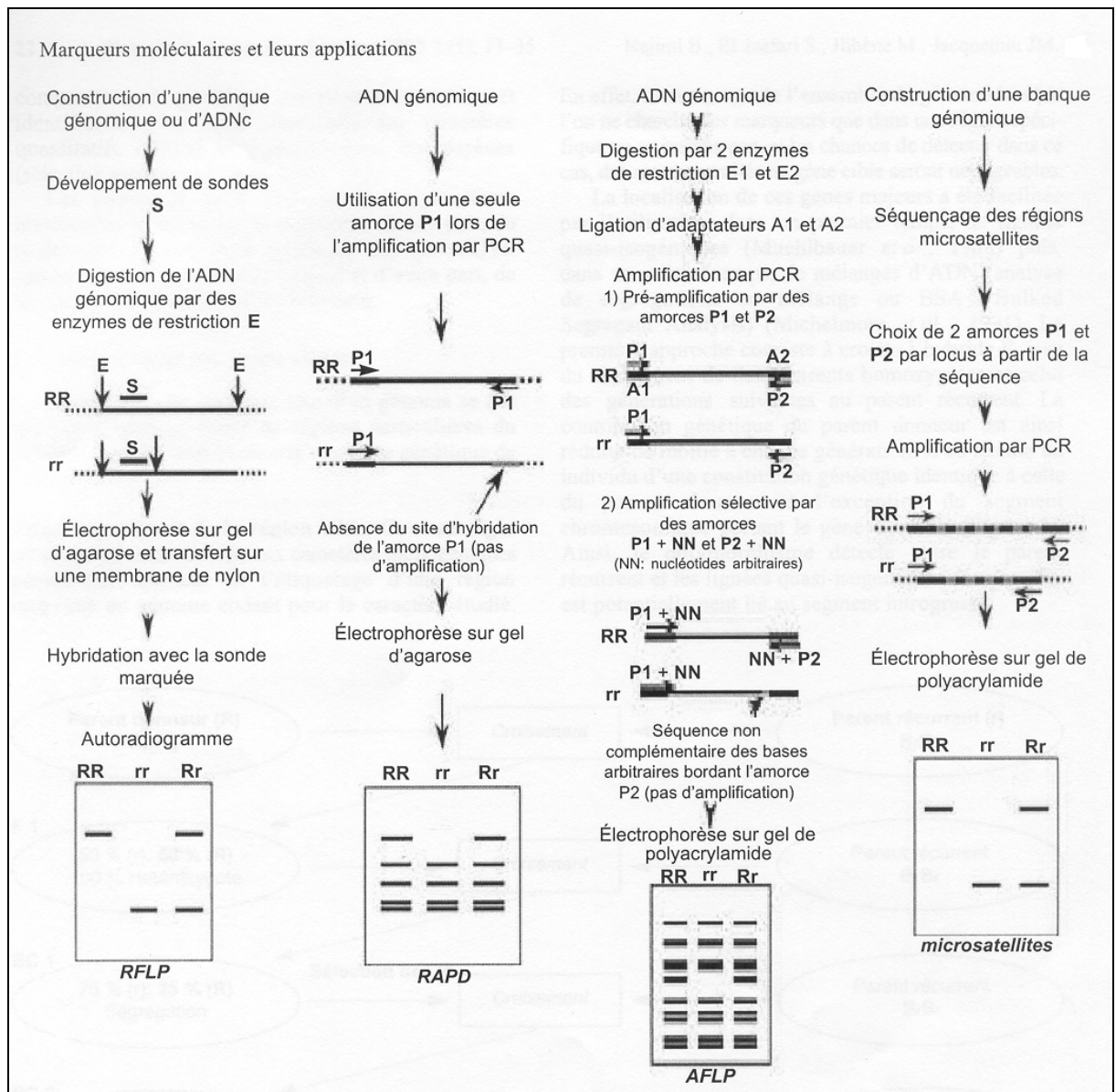


Fig. 9. Comparaison des principales techniques de marquage moléculaire (Bouchra Najimi *et al.* 2003).

2.2.3. Tableau récapitulatif

	RFLP	microsatellite	AFLP	RAPD
Type de marqueurs	Enzyme/sonde	Amorces spécifiques bordant les microsatellites	Enzyme/amorces	Amorces de séquences arbitraires
Nombre de marqueurs	Illimité	Quelques milliers	Illimité	Illimité
Polymorphisme révélé	Elevé	Très élevé	Très élevé	Très élevé
Neutralité	Oui	Oui	Oui	Oui
Codominance (1)	Oui	Oui	Non	Non
Spécificité du locus	Oui	Oui	Non	Non
Effort de développement	Elevé	Elevé	Moyen	Faible
Effort de génotypage	Elevé	Faible	Très faible	Très faible
Reproductibilité	Parfaite	Parfaite	Elevée	Faible
Principale utilisation (2)	Cartographie, génétique quantitative	Cartographie, génétique quantitative	Clonage positionnel, rétrocroisement	Clonage positionnel, rétrocroisement

La figure 9 reprend les différentes techniques utilisées pour mettre en évidence les marqueurs RFLP et les marqueurs utilisant la PCR.

(1) les marqueurs RFLP et microsatellite sont **codominants**. En effet, il est possible de distinguer sur un gel, un hétérozygote qui possédera deux fragments de tailles différentes d'un homozygote qui n'en aura qu'une.

Les marqueurs RAPD et AFLP sont des marqueurs mettant en évidence un polymorphisme de séquence. Si une mutation est survenue dans le site de restriction pour les marqueurs AFLP ou dans le site d'hybridation pour les marqueurs RAPD, alors soit l'amplification de ce locus n'a pas eu lieu et aucune bande n'est visible pour cet allèle, soit l'amplicon est plus grand que celui de l'autre allèle et il est impossible de savoir où celui-ci se trouve sur le gel. Ainsi, peu importe que l'individu soit homozygote ou hétérozygote pour un allèle donné ; il n'y aura jamais qu'une seule bande correspondant à celui-ci. Ce sont des marqueurs **dominants**.

(2) les marqueurs RFLP et microsatellites sont spécifiques d'un locus. Par ce fait, ceux-ci sont plus adaptés pour la mise en évidence de loci associés à un caractère quantitatif (QTL) ainsi que pour la cartographie.

Les marqueurs AFLP et RAPD ne sont pas spécifiques de locus. Ceux-ci sont donc plus utilisés pour la saturation d'une région particulière du génome permettant ainsi de densifier le marquage au voisinage d'un gène d'intérêt. Le clonage positionnel ainsi que le rétrocroisement n'en seront que facilités.

Les principales applications des marqueurs moléculaires sont décrites brièvement dans le paragraphe suivant.

2.3. Application des marqueurs moléculaires

2.3.1. Empreinte génétique

a. Principe

L'empreinte génétique est caractérisée par une succession de bandes correspondant à des fragments d'ADN après application d'une technique de mise en évidence du polymorphisme. Il est possible de prouver que deux individus sont génétiquement distincts en analysant le profil des bandes d'ADN sur un gel d'électrophorèse.

b. Applications

Les applications de l'empreinte génétique peuvent être réalisées à différents niveaux pour les végétaux :

- Identification variétale : permet de distinguer les lignées et de reconnaître les parents hybrides.
- Contrôle de la pureté : a pour but de détecter les contaminations variétales des lots de semence ou de tester l'homogénéité d'une population à n'importe quel stade d'un schéma de sélection.
- Protection variétale : fait partie intégrante de l'identification variétale. Il devient possible de détecter les contrefaçons ou copies de génotypes.

2.3.2. Cartes génétiques

a. Cartographie des marqueurs moléculaires

Les premières cartes génétiques partielles du maïs furent publiées en 1935, et réalisées à l'aide de marqueurs morphologiques. Les cartes génétiques permettent de représenter la disposition des gènes ou des marqueurs sur un chromosome. Elles se basent sur le principe de ségrégation des gènes.

Si deux gènes ne se trouvent pas sur un même chromosome, alors il ne ségrégeront pas ensemble lors de la méiose. Un croisement entre un hétérozygote et un homozygote donnerait une descendance de ce type :

		AaBb x AABB	
et		AB	
	AB	AABB	il y a donc 50% de génotypes parentaux
	Ab	AABb	
	aB	AaBB	50% de génotypes recombinants.
	ab	AaBb	

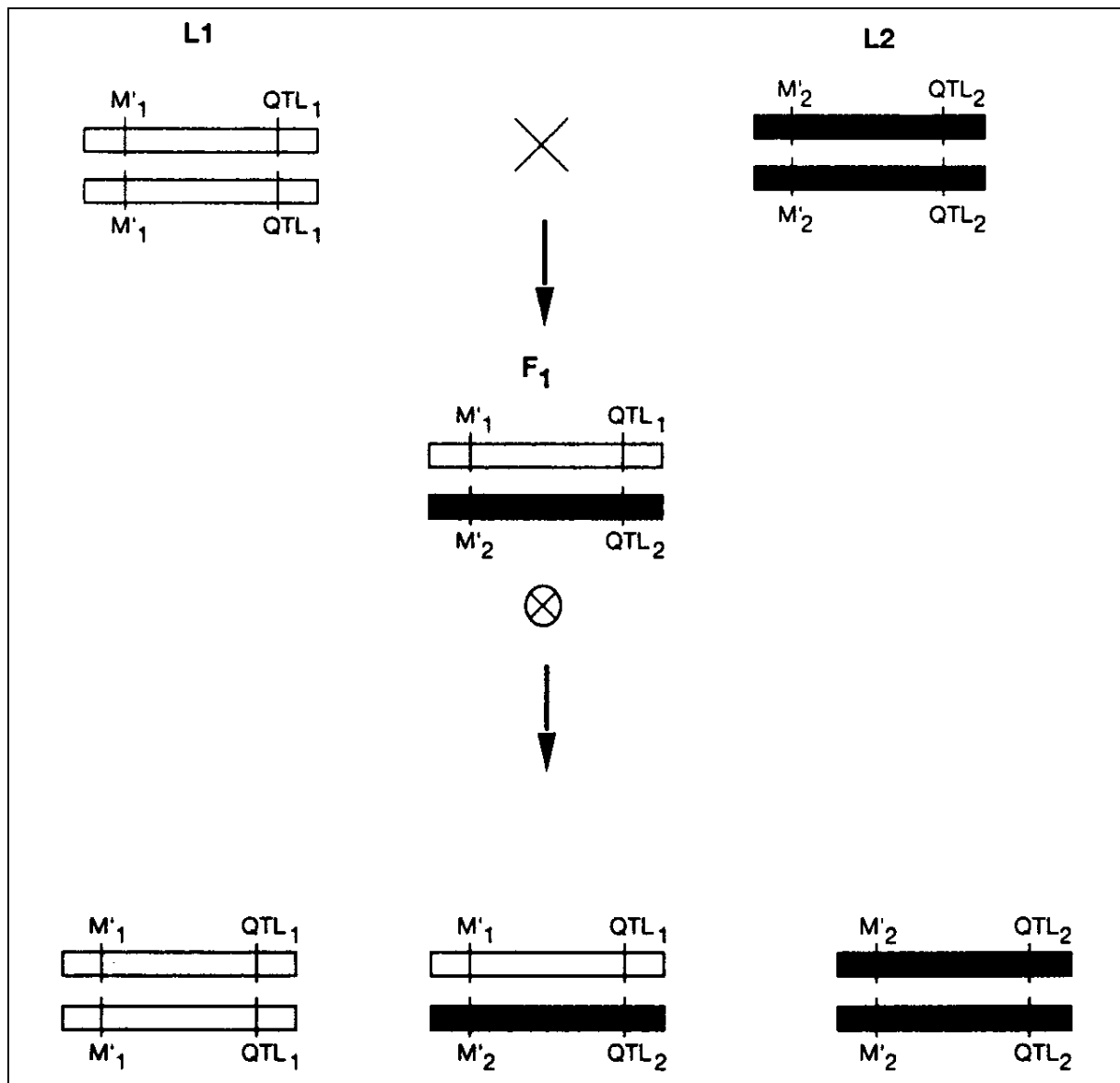


Fig. 10. Mise en évidence d'un QTL. Pour chaque génotype (M'₁M'₁, M'₁M'₂ ou M'₂M'₂) d'un marqueur donné à la génération F₂, on associe une valeur d'un caractère quantitatif. Si les valeurs sont significativement différentes entre les différents groupes, alors on peut associer ce marqueur au caractère quantitatif. Ici, dans ce modèle à un seul QTL, le loci QTL (QTL 1 et QTL 2) est associé au loci marqueur (M' 1 et M' 2). A partir du génotype du marqueur, il sera donc possible de connaître le génotype du QTL des individus.

Par contre, si deux gènes se retrouvent sur le même chromosome, alors on aura une descendance tout à fait différente :

$$AaBb \times AABB$$

	AB
AB	AABB
Ab	AaBb

On pourrait donc s'attendre à obtenir 100% de génotypes parentaux, mais ce n'est pas le cas. En effet, pendant la méiose, les chromosomes homologues subissent des crossing-over, ce qui entraîne un certain pourcentage de génotypes recombinants. Ce pourcentage est proportionnel au nombre de crossing-over qui lui-même est proportionnel à la distance entre les gènes.

Il est donc possible de savoir quels gènes sont liés et d'estimer la distance entre ceux-ci.

b. Cartographie d'un gène majeur

Quand un caractère est dû à l'effet d'un gène principal, on parle de gène majeur. L'apparition des marqueurs moléculaires a permis l'élaboration de méthodologies pour localiser ces gènes. Le principe est le même que celui servant à l'élaboration d'une carte génétique mais la différence est que l'on n'étudie plus la ségrégation de deux marqueurs mais d'un marqueur et d'un gène majeur.

Une fois la liaison établie entre ceux-ci, il est possible de connaître le phénotype futur de l'individu par détermination de la présence ou non du marqueur et cela avant le développement total de la plante.

Ce type de cartographie permet de réaliser du clonage positionnel. Le clonage positionnel d'un gène consiste à isoler le gène à partir des marqueurs qui lui sont liés. Pour cela, il est nécessaire de disposer de marqueurs se trouvant à la distance physique la plus faible possible du gène recherché. Les différents fragments d'ADN obtenus sont insérés dans des plasmides. On obtient des colonies de bactéries ou clones, ayant recombiné dans leur plasmide des fragments d'ADN se chevauchant partiellement. Par leur criblage à l'aide de marqueurs, on espère « atterrir » directement sur les clones contenant le gène.

c. Cartographie d'un caractère quantitatif

De nombreux caractères sont des caractères mesurables à variation continue comme le rendement, la précocité, la taille, la qualité des fruits. Dans ce cas il n'y a plus de distinction tranchée entre deux phénotypes comme par exemple résistance/sensibilité. On admet que plusieurs gènes sont impliqués dans le contrôle de ces caractères dits quantitatifs, et que de nombreux allèles sont responsables de la variabilité. Ces locus sont appelés QTL (Quantitative Trait Loci ; Locus de Caractères Quantitatifs).

Prenons un exemple où l'on croise deux individus différents pour un caractère quantitatif et considérons un marqueur quelconque présentant un polymorphisme chez ces deux parents. Dans la génération F2, il existe trois génotypes pour ce marqueur. Un quart de chaque homozygote et la moitié de plantes hétérozygotes. (Fig. 10.)

L'idée est de réaliser trois classes d'individus sur la base du génotype de ce marqueur, et de comparer les moyennes du caractère quantitatif des individus de ces trois classes.

Si les moyennes sont déclarées significativement différentes, cela signifie que, sur la base du génotype du marqueur, il est possible de différencier trois classes d'individus pour les QTL.

La première application est donc de prédire la valeur génétique de la descendance d'un croisement. On évalue donc plus précisément la valeur des individus candidats à la sélection.

Une autre application repose sur la mise en évidence d'allèles favorables aux QTL. Il est alors possible de cumuler ces allèles chez un individu par rétrocroisement notamment. On parle alors de construction de génotypes.

2.3.3. Sélection assistée par marqueurs

Les rétrocroisements sont utilisés pour réaliser l'introgression d'un gène dans une variété élite. Cette opération est appelée également conversion. Mais en dépit du nombre de rétrocroisements, il reste toujours dans la lignée receveuse un plus ou moins grand fragment du parent donneur autour du gène d'intérêt.

Les marqueurs moléculaires sont souvent utilisés pour conduire la conversion, car ils permettent de réduire les temps de sélection. En effet, on peut choisir à chaque génération les plantes ayant recombiné le plus petit segment chromosomique, c'est-à-dire les plantes possédant uniquement les marqueurs les plus proches du gène d'intérêt.

(C. Dagorn, 1999; D. de Vienne, 1998)

3. Les microsatellites

Une compréhension complète des mécanismes régulant les microsatellites est essentielle pour assurer l'utilisation correcte de ces marqueurs comme outils analytiques.

3.1. Définition

Plusieurs termes définissent les séquences répétées présentes dans le génome. Certains ne sont pas encore correctement définis et parfois même, possèdent des définitions différentes suivant les auteurs. Tautz (1993) (*in* G. K. Chambers *et al.*, 2000) définit ces différents termes comme ceci :

- **Satellites** : segments répétés de 100 nucléotides ou plus formant un fragment plus ou moins uniforme de 10^3 - 10^7 nucléotides de long.
-
- **Minisatellites** : segments répétés de 10-100 nucléotides formant un fragment plus ou moins uniforme de 10^2 - 10^5 nucléotides de long.
-
- **Microsatellites** : segments répétés de 2 à 6 nucléotides formant un fragment plus ou moins uniforme d'environ 10^2 nucléotides de long.

- **Fragments mononucléotidiques** : série uniforme de mononucléotides de n'importe quelle taille.

Il est important de faire une différence entre les minisatellites et les microsatellites, même si la taille de leurs segments répétés ne diffère parfois que de quelques nucléotides.

En effet, ces deux répétitions de séquence diffèrent par leurs mécanismes mutationnels. La variation du nombre de répétitions dans un microsatellite est due au mauvais appariement d'un des deux brins d'ADN lors de la réplication (voir point 3.4). Quand au développement des minisatellites, il semblerait provenir d'un mécanisme polaire (spécifique de la région 3') incluant un échange des régions bordantes. Malheureusement, le mécanisme mutationnel des éléments répétés comprenant des unités de taille entre 7 et 10 nt est très peu connu.

3.2. Différence de distribution des microsatellites entre les régions codantes et non codantes

Les microsatellites sont largement représentés dans l'ADN non codant et sont relativement rares dans les régions codantes. Cette affirmation est soutenue par plusieurs observations :

- Tous les types de SSR (de la répétition mono- à hexanucléotidique) ont été trouvés en excès (en comparaison à une distribution aléatoire) dans les régions non codantes du génomes de différents organismes : *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Shizosaccharomyces pombe*, *Mus musculus*, *Drosophila melanogaster*, plantes et primates (Metzgar *et al.* 2000 in Y. Li *et al.*, 2002).
- Morgan *et al.* (2002) (in Y. Li *et al.*, 2002) rapportent que chez six espèces de plantes incluant Arabidopsis, le riz, le maïs, le soya, et le froment, tous les microsatellites, excepté les triplets et les hexonucléotides, sont significativement moins fréquents dans les séquences codantes prédites comparées à la fraction non codante du génome.
- Dans le génome du poisson japonais *Fugu rubripes*, seulement 11% des 6042 SSR ont été détectés dans la partie codante du génome (Edwards *et al.* 1998 in Y. Li *et al.*, 2002).

Au niveau des régions codantes, des microsatellites dinucléotides ont été observés dans les régions 3' et 5' non transcrites des gènes ainsi que dans les introns.

Malgré ces constatations, les microsatellites se font beaucoup plus rares dans ces régions. En effet, l'expansion de la taille des SSR dans les régions 5' et 3' ainsi que dans les introns pourrait conduire à une perturbation de la protéine originale et/ou à la formation d'un nouveau gène par décalage de la phase de lecture (les éléments répétés dans les introns joueraient un rôle dans l'épissage). Cela suggère donc que la distribution aléatoire est sévèrement contre-sélectionnée (Bachtrog *et al.* 1999 ; Liu *et al.* 1999 in Y. Li *et al.*, 2002), ce qui explique la différence de distribution des microsatellites entre les régions codantes et non codantes. En effet, la sélection va éliminer les mutations décalant le cadre de lecture dans les régions codantes ; c'est-à-dire les variations n'étant pas un multiple de trois (Liu *et al.* 1999 ; Dokholyan *et al.* 2000 in Y. Li *et al.*, 2002). Cette explication est confirmée par le fait que la distribution de SSR trinuécléotidiques est identique pour les régions codantes et non codantes du génome des champignons, des protistes, des procaryotes, des virus, des organites, des plasmides et des humains (Field & Wills 1996 ; Wren *et al.* 2000 in Y. Li *et al.*, 2002).

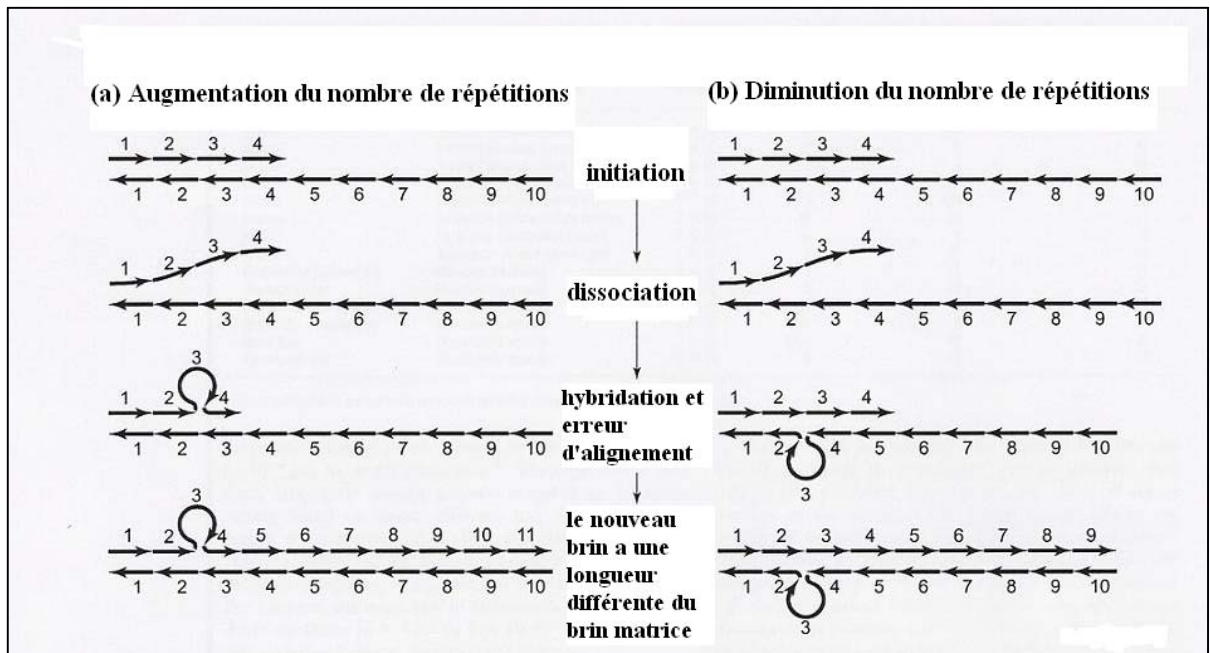


Fig. 11. Modèle de « slippage replication » permettant d'expliquer l'augmentation ou la diminution du nombre de répétitions dans un microsatellite (inspiré de l'article de H. Ellegren, 2000)

3.3. Conservation des marqueurs microsatellites entre les espèces

Si les régions bordant des microsatellites ne sont pas conservées entre le taxon à partir duquel les marqueurs ont été isolés et un autre taxon, alors ces amorces ne peuvent être utilisées chez celui-ci. La conservation de ces régions bordantes est directement proportionnelle au temps de divergence entre les espèces.

Quarante-six marqueurs isolés à partir d'ADN de bovin ont été étudiés chez le mouton, le cheval et l'homme. 56% des amorces sont conservées chez le mouton qui diverge du bovin d'environ 12 à 25 millions d'années. Seuls 5,6% des marqueurs sont amplifiés chez le cheval et aucune amplification n'a été observée chez l'Homme qui est séparé des bovins depuis environ 80 à 100 millions d'années.

Certains marqueurs sont conservés entre des tortues marines divergeant par plus de 300 mille ans. Cette conservation s'expliquerait par le fait que le taux d'évolution moléculaire serait plus lent chez cette espèce.

Chez les astéracées, famille à laquelle appartient la chicorée industrielle, le maximum de divergence entre deux taxons pour lequel il est possible d'avoir une conservation des amorces est estimé entre 14,5 et 29 millions d'années.

La conservation des régions bordantes entre les différentes espèces d'astéracées est donc plutôt faible. Cette observation implique que la vitesse d'évolution entre les différentes espèces d'astéracées est élevée (J. Whitton *et al.*, 1997).

3.4. Modèles et taux de mutation des microsatellites

Ce point se base sur l'article de synthèse de H. Ellegren (2000)

3.4.1. Modèles

Deux hypothèses ont été proposées afin d'expliquer l'évolution du nombre de répétitions dans un microsatellite. Le premier modèle, appelé « **slippage replication** », a été décrit par Levinson et Gutman il y a environ 15 ans (Fig. 11).

Ce modèle est composé de trois étapes :

- ***étape d'initiation*** : la polymérase synthétise le brin complémentaire de la séquence répétée.
- ***étape de dissociation*** : le brin nouvellement synthétisé se dissocie du brin matrice pendant que la polymérase traverse la région répétée.
- ***étape de rehybridation et erreur d'alignement*** : deux cas sont possibles :
 - Ajout d'une unité de répétition :

L'extrémité 3' du brin nouvellement synthétisé se réhybride avec une région complémentaire en amont du brin matrice. Il y a ainsi formation d'une boucle dans le nouveau brin. La polymérase va donc ajouter une unité de répétition supplémentaire.

- Suppression d'une unité de répétition :

Ce mécanisme est presque identique à l'ajout d'une unité avec la différence que la boucle se fait au niveau du brin matrice. La polymérase synthétisera donc une répétition en moins.

Le deuxième modèle de mutation est le **crossing-over inégal** pendant la méiose.

Certaines preuves clés montrent que le modèle principal est celui du « slippage replication ». En effet :

- L'analyse des différents allèles pour un locus microsatellite particulier montre que les régions bordantes de celui-ci ne subissent pas de phénomène de recombinaison. Cela suggère que le phénomène de mutation est principalement intra-allélique et donc qu'il n'y a pas d'échange entre les différents allèles.
- Le taux et le profil de mutation semblent ne pas changer entre les chromosomes hemizygotes (le chromosome Y) et les chromosomes pairés (autosomes chez les diploïdes). Cela suggère que la mutation ne requiert pas de contact entre les chromosomes homologues.
- *In vitro*, certaines expériences ont montré clairement que les séquences microsatellites ont une capacité intrinsèque à subir un « DNA slippage ».

3.4.2. Taux de mutation

Ces mécanismes particuliers associés aux SSR vont influencer le taux de mutation. Ce taux est approximativement égal à 10^{-9} mutations par nucléotide et par génération pour une séquence unique dans l'ADN eucaryote.

Le taux de mutation chez les microsatellites, lui, est de l'ordre de 10^{-3} - 10^{-4} par locus et par génération. Celui-ci varie d'une espèce à l'autre.

Le taux de mutation des SSR ne diffère pas uniquement entre les espèces mais aussi au sein même d'une espèce. En effet, les longs loci mutent plus rapidement que les courts. La longueur des microsatellites influence donc le taux de mutation.

Dans le modèle classique de mutation décrit ci-dessus, certaines modifications doivent être apportées :

- Par mutation, une seule répétition est enlevée ou ajoutée, mais il semblerait que plusieurs répétitions peuvent être ajoutées ou enlevées en même temps.
- On estime que la probabilité d'une addition ou d'une délétion est identique et constante à travers les allèles. Cependant, la mutation semblerait plus en faveur de l'expansion de la région répétée. Cette observation a été faite lors d'études sur les loci téranucléotidiques ainsi que sur les loci humains dinucléotidiques.

3.5. Cycle de vie d'un microsatellite

Le point 3.5. se base sur les articles de synthèse suivants : Y. Li *et al.* (2002) ; G. K. Chambers *et al.* (2000) et H. Ellegren (2000).

Les microsatellites sont souvent visualisés comme ayant un « cycle de vie ». En effet, ces séquences naissent, grandissent et finalement « meurent ». Éventuellement, ces fragments peuvent même ressusciter ! Ces événements peuvent prendre une dizaine, voire une centaine de millions d'années

3.5.1. Naissance

Les microsatellites naissent dans une région dite d'une « simplicité cryptique ». Le terme « simplicité cryptique » définit une courte séquence (< 8 nt de long) composée de répétitions. Pour que le développement du microsatellite puisse commencer, il faut que les répétitions successives soient au minimum huit. Il faut donc que des mutations ponctuelles relient des régions cryptiques entre elles.

Par exemple, imaginons une séquence contenant une région cryptique : (GT)₃-(AT)-(GT)₃ ; la région cryptique est représentée par la triple répétition de GT. Si une mutation ponctuelle remplace AT en GT alors huit répétitions se suivront. Cette mutation correspond à la naissance du microsatellite.

3.5.2. Croissance

Les répétitions en tandem vont subir des mutations. Une bonne partie de ces mutations va être corrigée par le système de réparation cellulaire. Mais de temps en temps, ces mutations ne sont pas réparées. La grande majorité de celles-ci sont des gains de répétitions. Notre microsatellite entame donc sa croissance.

Au fur et à mesure que le microsatellite croît, les mutations sont de plus en plus fréquentes car le taux de mutation est proportionnel à la longueur du microsatellite. La vitesse de croissance augmente. On obtient un microsatellite dit « parfait ».

3.5.3. Mort et résurrection

Il semblerait que le processus de croissance d'un microsatellite ne soit pas infini. Les microsatellites peuvent être frappés par des mutations ponctuelles. Si l'allèle parent est une répétition parfaite, l'allèle résultant contiendra une interruption. La répétition sera donc imparfaite car séparée par un nucléotide quelconque. Les répétitions imparfaites sont moins variables que les SSR parfaits. Cela implique une réduction du taux de mutation.

La taille maximale d'un microsatellite serait donc déterminée par un équilibre entre les mutations apportant des répétitions supplémentaires et les mutations coupant une répétition en tandem parfaite en deux répétitions de séquences imparfaites

Ainsi, l'évolution des microsatellites est un processus où le slippage conduit arbitrairement certains allèles à la croissance en taille jusqu'à ce que les mutations ponctuelles coupent ces répétitions parfaites.

A ce moment, l'expansion des deux nouveaux microsatellites va ralentir. Soit la croissance redémarre et le cycle recommence. Soit les mutations ponctuelles s'accumulent jusqu'à obtenir un nombre insuffisant de répétitions pour recommencer la croissance. Dans ce dernier cas, le microsatellite est mort.

Après la mort du microsatellite, la résurrection est envisageable. En effet, le cycle peut recommencer grâce à des mutations supprimant les interruptions entre les séquences répétées, donnant lieu à 8 répétitions successives.

3.6. Exemples d'applications des marqueurs microsatellites

3.6.1. Création de groupe de liaison

Le principe de la cartographie à l'aide de marqueurs moléculaires a été expliqué précédemment au point 2.3.2.

Ce type d'étude a par exemple été réalisé sur *Meleagris gallopavo* (dinde). Les groupes de liaison de la poule (*Gallus gallus*) ont déjà été déterminés. Comme ces deux espèces sont proches, les chercheurs de l'Institut Roslin à Edimbourg en Ecosse ont pensé utiliser les microsatellites isolés chez la poule pour cartographier le génome de la dinde. Malheureusement, seulement 10 à 20% des microsatellites de la poule étaient polymorphes chez la dinde. Ils se sont donc résignés à créer leurs propres microsatellites enrichis en GT. Au total, 111 marqueurs polymorphes répartis dans 20 groupes de liaison ont été obtenus. Cette carte recouvre 700 cM du génome, soit 25% de la totalité. D'autres marqueurs doivent donc être développés pour compléter cette carte.

Grâce à ce travail, l'identification de QTL ainsi que la comparaison de cartes par comparaison des marqueurs d'autres espèces sont maintenant possibles chez la dinde (D.W Burt *et al.* 2003).

3.6.2. Utilisation des microsatellites pour la traçabilité dans la filière animale

La traçabilité est définie comme étant l'aptitude à retrouver l'historique, l'utilisation ou la localisation d'un produit au moyen d'identifications enregistrées. Avant l'utilisation de microsatellites dans la filière animale, les animaux étaient reconnus soit à l'aide de descriptions, de marques déposées sur l'animal, de marquages électroniques ou encore de marquages phénotypiques basés sur les groupes sanguins. L'utilisation de microsatellites a permis de supprimer tous les désavantages des méthodes précédentes. Cette technique a donc été développée et l'empreinte génétique peut maintenant être associée aux documents administratifs de l'animal. Cela permet d'assurer la traçabilité moléculaire des animaux et de leurs produits tout au long d'une filière de production (D. Portetelle *et al.* 2000).

3.6.3. Utilisation des microsatellites pour la détermination de l'introggression au sein d'une espèce végétale

L'introggression est définie comme étant la dispersion naturelle des gènes d'une espèce à l'intérieur d'une autre espèce par hybridation interspécifique suivie de plusieurs rétrocroisements avec le parent local. L'introggression peut amener à une pollution génétique c'est-à-dire une contamination de bassins génétiques de populations ou d'espèces indigènes par du matériel génétique provenant de variétés domestiquées, d'organismes génétiquement modifiés (OGM), d'espèces envahissantes,... Ces nouvelles combinaisons de gènes pourraient avoir des conséquences négatives pour les populations d'organismes adaptées aux conditions environnementales locales. Bien que les incidences écologiques ne soient pas entièrement comprises, elles dépendent sans aucun doute de l'importance et de la fréquence de la contamination des populations indigènes par des gènes non indigènes.

L'introgression de matériel génétique non indigène peut être surveillée par l'utilisation d'outils génétiques (par exemple les microsatellites) facilitant l'étude et la quantification de l'effet d'une telle population génétique sur les espèces indigènes.

Ce genre d'étude a été réalisé par une équipe de chercheurs italiens sur le peuplier noir européen (*Populus nigra* L.). A l'aide de six microsatellites, ils ont pu déterminer, en réalisant l'empreinte génétique de plusieurs espèces sauvages et domestiques présentes dans une même région, qu'il y avait introgression entre l'espèce naturelle et les plantations hybrides monoclonales cultivées dans les environs. En effet, certains allèles spécifiques de l'espèce domestique se retrouvaient dans l'espèce sauvage (T. Fossati *et al.* 2003).

3.6.4. Exemple d'utilisation des microsatellites pour l'identification de QTL

Pour rappel, la cartographie des QTL consiste à associer un caractère quantitatif avec un marqueur génétique. Ainsi, le génotype du marqueur renseignera, bien avant que l'on puisse l'observer, sur le phénotype du caractère quantitatif qui nous intéresse. Les microsatellites sont souvent requis pour ce genre d'utilisation.

Des chercheurs américains ont ainsi développé 181 microsatellites dans le but de cartographier des QTL liés à la résistance au froid chez le riz. Leur démarche a été la suivante : tout d'abord ils ont croisé une variété de riz résistante au froid avec une autre variété ne possédant pas cette résistance. Le but de cette étape est de créer une population recombinante possédant les gènes de résistance au froid. Les hybrides ont été soumis à des conditions extrêmes de froid (4°C) et les individus résistants ont été sélectionnés. La dernière étape fut de trouver les loci microsatellites dont le génotype est corrélé avec le phénotype étudié. Plusieurs QTL associés à la résistance au froid ont ainsi pu être déterminés (V.C. Andaya *et al.* 2003).

3.6.5. Les microsatellites en génétique des population

La génétique des populations étudie la variabilité ou diversité génétique au sein d'individus en interaction avec leur environnement.

L'avènement des marqueurs moléculaires a permis de développer cette discipline. En effet, l'accès au niveau même de l'ADN a permis de découvrir une hétérogénéité considérable du génome en ce qui concerne le polymorphisme et le mode d'évolution (D. de Vienne, 1998).

Pour chaque marqueur, on peut donc trouver un certain nombre d'allèles (pour un locus microsatellite, un allèle est caractérisé par un nombre particulier de répétitions) à partir desquels on pourra calculer la fréquence allélique ainsi que le taux d'hétérozygotie pour une population donnée (Fig.12.a).

Ensuite, il est possible de calculer le taux de subdivision de la population au moyen des F-statistiques définies par Wright (1943) (Fig.12.b). Ces statistiques correspondent aux corrélations d'allèles d'un même locus pris à un certain niveau de subdivision par rapport à des allèles d'un même locus pris au niveau supérieur de subdivision :

- F_{it} représente la corrélation de deux allèles d'un individu (i) par rapport à deux allèles pris au hasard dans la population totale (t).
- F_{is} représente la corrélation de deux allèles d'un individu (i) par rapport à deux allèles pris au hasard dans une sous-population (s). Cette mesure est aussi appelée coefficient de consanguinité. Il met en évidence la diminution ou l'augmentation du taux

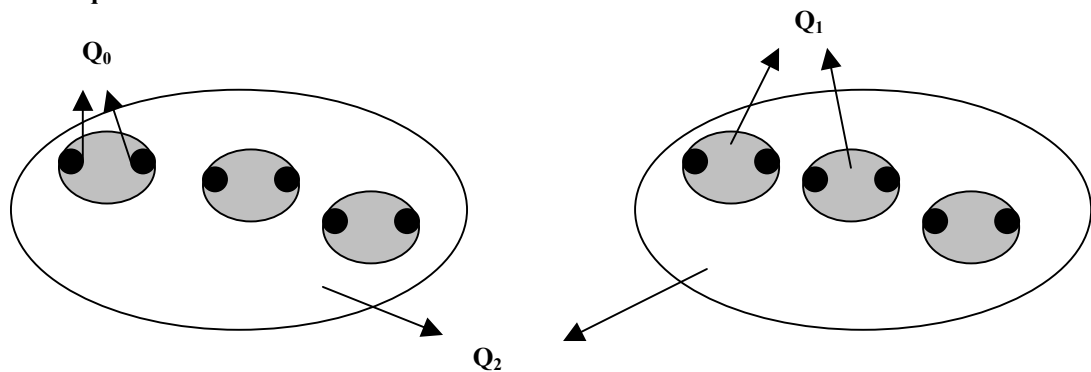
a) soit un locus biallélique A

Génotype de l'échantillon :
 AA : X individus
 Aa : Y individus
 aa : Z individus

Fréquence allélique de A : $\frac{2X+Y}{2(X+Y+Z)} = \frac{\text{Nombre d'allèles A}}{\text{Nombre total d'allèles}}$

Taux d'hétérozygotie : $\frac{Y}{X+Z} = \frac{\text{Nombre d'hétérozygotes pour un locus}}{\text{Nombre d'homozygotes}}$

b) F-Statistiques :



Cercle blanc : sous population
 Cercle gris : individus
 Cercle noir : allèles

Q = probabilité que 2 allèles d'un même locus soient identiques.

Q₀ = 2 allèles d'un même locus chez un même individu.

Q₁ = 2 allèles d'un même locus d'individus différents dans une même sous-population.

Q₂ = 2 allèles d'un même locus de sous-populations différentes.

i = individu ; s = sous-population ; t = population totale

$$F_{it} = \frac{Q_0 - Q_2}{1 - Q_2} \approx \frac{i}{t}$$

$$F_{is} = \frac{Q_0 - Q_1}{1 - Q_1} \approx \frac{i}{s}$$

$$F_{st} = \frac{Q_1 - Q_2}{1 - Q_2} \approx \frac{s}{t}$$

0-0,05 : petite différence génétique
 0,05-0,15 : différence génétique modérée
 0,15-0,25 : grande différence génétique
 > 0,25 : très grande différence génétique

Autre formule pour F_{st} : $\frac{1}{1+4N_e(m+\mu)}$

car la variation entre populations est due à
 - la dérive (N_e = taille de la population)
 - au flux de gènes (m)
 - aux mutations (μ)

Fig. 12. a) Calcul des fréquences alléliques et des taux d'hétérozygotie. b) formules des F-Statistiques et significations.

d'hétérozygotie correspondant respectivement à une augmentation ou à une diminution de variabilité au sein de la sous population.

- F_{st} représente la corrélation entre deux allèles pris dans une sous-population par rapport à deux allèles pris au hasard dans la population totale.

F_{is} et F_{it} sont calculés à partir des taux d'hétérozygotie, car on travaille à l'échelle de l'individu. Par contre, le F_{st} est calculé à partir des fréquences alléliques, car on est au niveau des sous-populations.

Avant de commencer un programme de sélection sur une espèce, il est bon de savoir si la variabilité génétique dans la population de base est suffisante afin que celle-ci puisse éventuellement faire face aux conditions environnementales. Dans ce but précis, une étude sur la variabilité génétique chez quatre lignées domestiques du poisson Tilapia du Nil a été réalisée à l'aide de marqueurs microsatellites. Le laboratoire hollandais de génétique animale de l'université de Wageningen a sélectionné une dizaine de marqueurs microsatellites à partir de la banque de donnée NCBI. A partir des données obtenues pour chaque marqueur après amplification, les F-statistiques ont été calculées.

- Le F_{is} montre une certaine diminution du taux d'hétérozygotie observé par rapport au taux d'hétérozygotie attendu aux seins des lignées. Ce phénomène serait probablement dû à la nature finie de la population et donc à l'accumulation de la consanguinité. En effet, même avec un croisement aléatoire, lorsque la taille effective de la population est faible, une diminution de l'hétérozygotie est parfois observée.
- Le F_{st} montre une différence génétique entre les lignées allant de modérée à grande.

Il a été démontré que la perte d'une des quatre lignées entraînerait une perte de diversité entre 14 et 57 %. Cette information suggère que chaque lignée doit être conservée afin de maintenir la quantité de diversité génétique présente dans la totalité du groupe (M.J.M. Rutten *et al.* 2004).

La génétique des populations permet donc d'obtenir des informations sur l'organisation génétique d'un groupe d'individus.

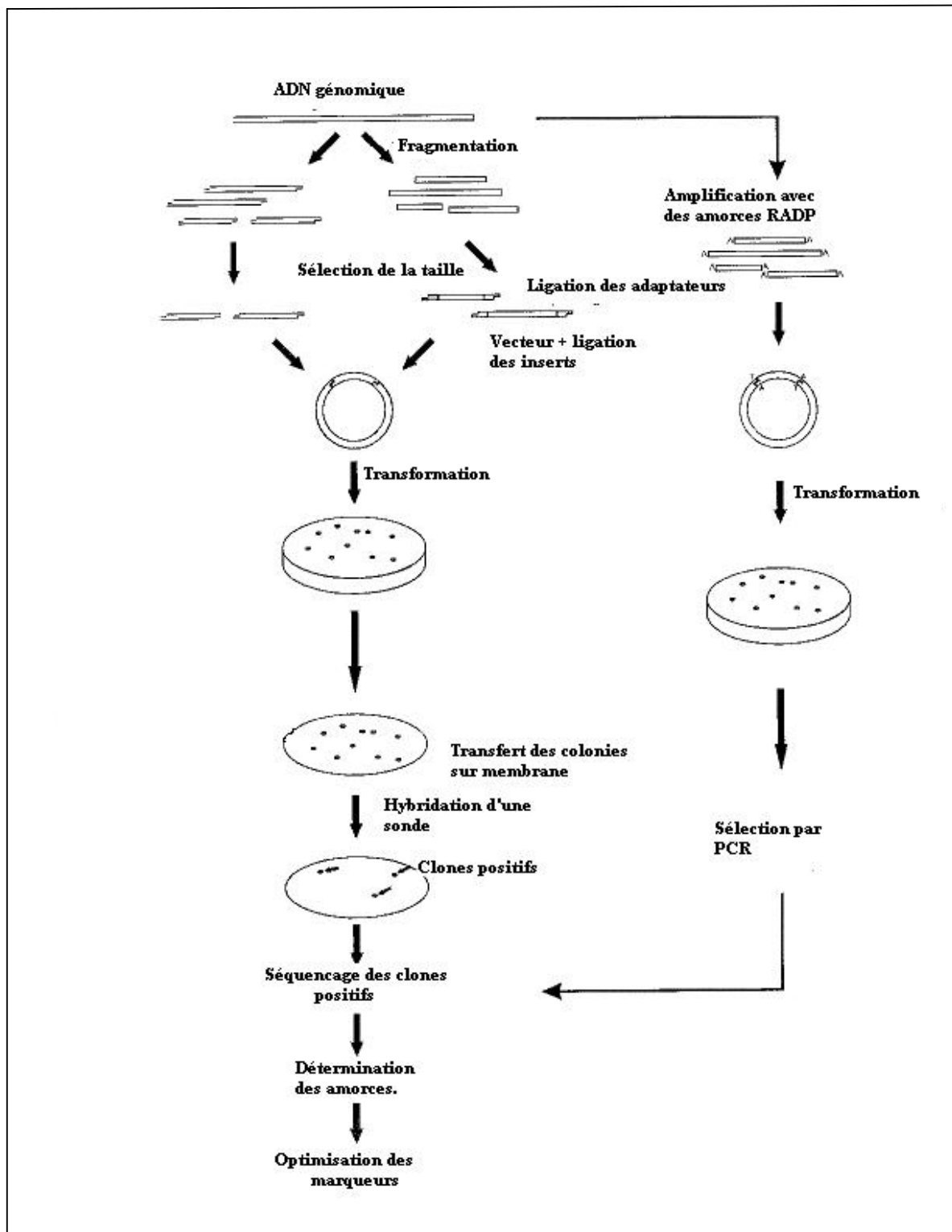


Fig. 13. La méthode classique consiste à générer une banque génomique et de détecter les clones positifs à l'aide d'une sonde. Une méthode alternative est de créer une banque génomique à partir d'amplicons obtenus avec des amorces RAPD. Les clones positifs sont détectés par PCR à l'aide d'amorces spécifiques du microsatellite. Pour les deux méthodes, les étapes suivantes sont le séquençage des clones, le design des amorces et l'optimisation des marqueurs.

4. Stratégie pour l'isolation des marqueurs moléculaires

Le défaut majeur des marqueurs microsatellites est la difficulté de les isoler. Plusieurs méthodes sont proposées (L. Zane *et al.*, 2002) :

4.1. Méthode originale

Historiquement, cette méthode est la première à avoir été utilisée. Elle consiste à prendre une banque génomique et à sélectionner les clones contenant un microsatellite à l'aide d'une sonde (Fig.13). Les principales étapes peuvent être définies comme suit :

- L'ADN génomique est fragmenté par des enzymes de restriction.
- Les fragments entre 300 et 700 pb sont sélectionnés.
- Ceux-ci sont intégrés dans un plasmide (directement ou après ligation d'adaptateurs).
- Les plasmides sont intégrés dans des bactéries.
- Les bactéries positives sont sélectionnées à l'aide d'une sonde (radioactive ou non) lors de l'hybridation sur membrane de nylon.
- Les amorces sont définies et synthétisées, les conditions PCR sont optimisées.

Au lieu de fragmenter l'ADN génomique avec des enzymes de restriction, certains chercheurs amplifient au hasard des fragments d'ADN avec des amorces RAPD. Ces fragments sont ensuite intégrés dans des plasmides. Les clones positifs sont détectés par PCR à l'aide d'amorces spécifiques du vecteur et des répétitions du microsatellite que l'on veut isoler. L'avantage principal dans l'utilisation d'amorces RAPD est qu'il semblerait que les microsatellites soient plus fréquents dans les fragments amplifiés que dans les clones générés au hasard.

Le pourcentage de clones positifs dans une banque génomique par la méthode traditionnelle tourne autour de 2-3%. Cette méthode demande donc beaucoup de travail pour un résultat plutôt faible.

4.2. Méthode utilisant l'extension d'amorces

Deux protocoles ont été proposés pour la réalisation d'une banque génomique enrichie en microsatellites par l'extension d'amorces (Ostander *et al.* 1992 et Paetkau 1999).

Les deux méthodes se basent sur la construction d'une banque génomique « primaire » (Fig. 14) : les fragments d'ADN sont insérés dans des phages ou des phagemides dans le but d'obtenir une librairie d'ADN génomique simple brin. Celui-ci sera utilisé comme matrice lors de la réaction d'extension des amorces spécifiques des répétitions que nous voulons isoler.

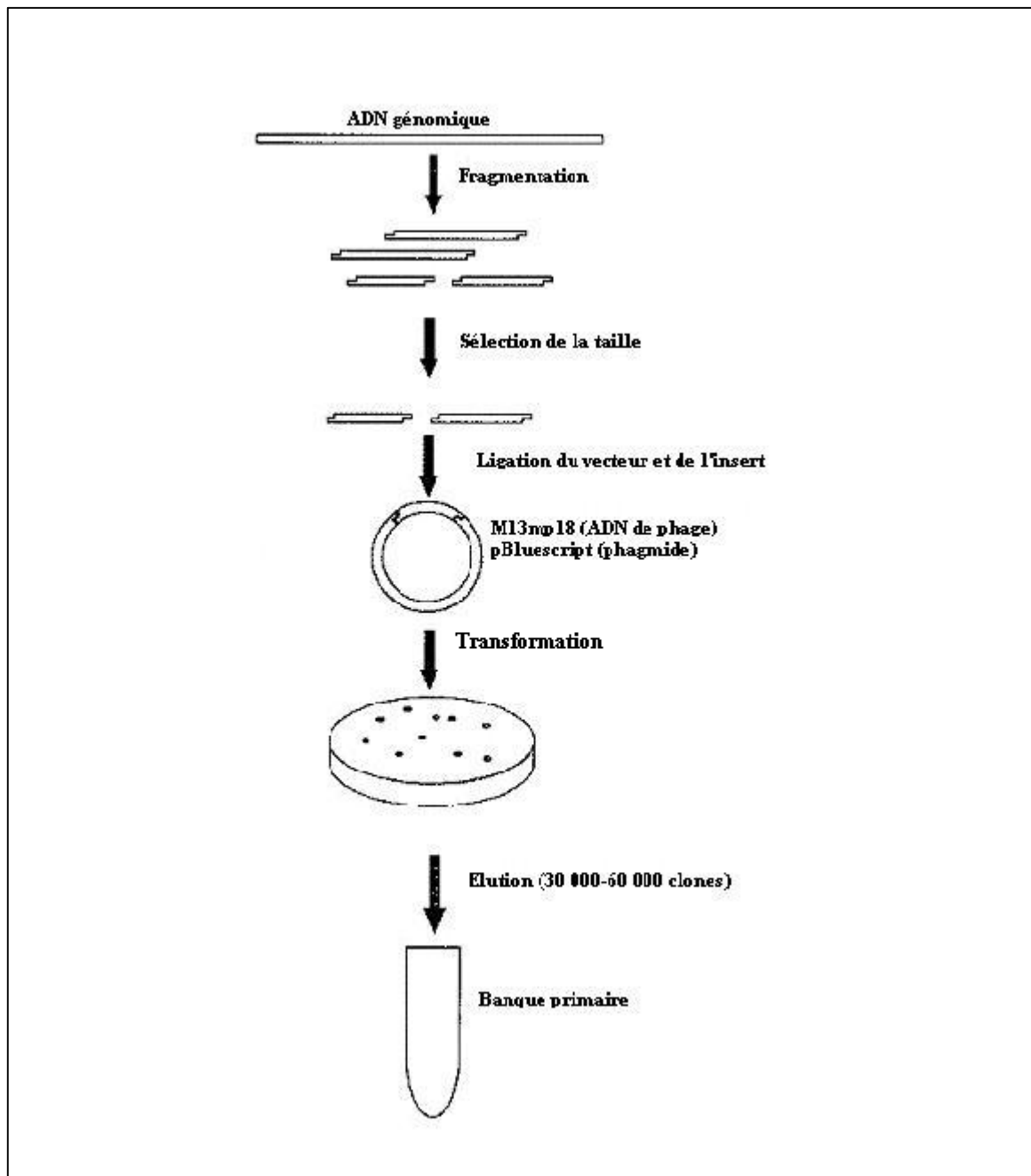


Fig. 14. Mise en place d'une banque primaire dans le but d'isoler des microsatellites. L'ADN génomique fragmenté est inséré dans un phage ou un phagmide puis inséré dans des bactéries. L'ADN simple brin du vecteur contenant les régions répétées sera détecté par PCR lors de l'amplification par des amorces spécifiques des répétitions.

4.4.1. Méthode de Ostender

Voici les différentes étapes décrites par Ostender *et al.* en 1992 (Fig. 15) :

- 40 000 à 60 000 colonies provenant de la banque de phagemides sont infectées par des phages helper M13. La bactérie hôte a un génotype particulier *dut-ung-*. *ung-* signifie que l'uracile N-glycosylase est inactivée. Cette enzyme reconnaît les uraciles dans l'ADN et clive la liaison N-glycosydique entre le désoxyribose et la base, formant ainsi un site apurinique. *dut-* signifie que l'activité de l'UTPase est inhibée. L'infection donnera donc de l'ADN circulaire simple brin contenant de l'uracile à la place de la thymine.
- Des amorces spécifiques des microsatellites d'intérêt sont ajoutées dans le milieu. Il y a conversion sélective de l'ADN simple brin contenant les répétitions en ADN double brin.
- Le mélange d'ADN double brin et simple brin est utilisé pour transformer des bactéries *dut+ung+*.
La banque qui en résulte est hautement enrichie en microsatellites car l'ADN simple brin transforme avec très peu d'efficacité et l'ADN simple brin contenant l'Uracile est dégradé par l'Uracil-N-glycosylase. Au total, seul l'ADN double brin est présent dans la bactérie. Ces bactéries constituent la banque secondaire.
- Les clones positifs sont ensuite sélectionnés par une sonde lors de l'hybridation sur membrane des colonies.

4.2.2. Méthode de Paetkau

Les différentes étapes décrites par Paetkau *et al.* en 1999 sont les suivantes (Fig.15) :

- 30 000 clones sont infectés avec le phage M13 permettant ainsi d'obtenir de l'ADN circulaire simple brin.
- La réaction d'allongement se fait à l'aide d'amorces biotinylées à l'extrémité 5'. Ce brin d'ADN va donc se lier avec des billes magnétiques couplées à la streptavidine.
- L'ADN simple brin obtenu après dénaturation et contenant les régions répétées sera utilisé pour la transformation des bactéries hôtes.
- Ces bactéries sont criblées par PCR et les clones positifs sont séquencés.

L'enrichissement par ces méthodes est vraiment efficace pour un enrichissement en répétitions AC : le pourcentage de clones positifs se situe entre 40 et 50 % pour la méthode décrite par Ostrander et jusqu'à 100 % pour la méthode de Paetkau. Malheureusement, ces techniques demandent un nombre élevé d'étapes et fonctionnent beaucoup moins bien pour les tri- et tétranucléotides.

4.3. Isolation par hybridation sélective

La méthode la plus simple semble être celle publiée par Hamilton en 1999. En voici les différentes étapes (Fig.16.) :

- Les premières étapes sont identiques au protocole traditionnel : l'ADN génomique est fragmenté par des enzymes de restriction ; ses fragments sont insérés dans un plasmide ou liés à des adaptateurs.

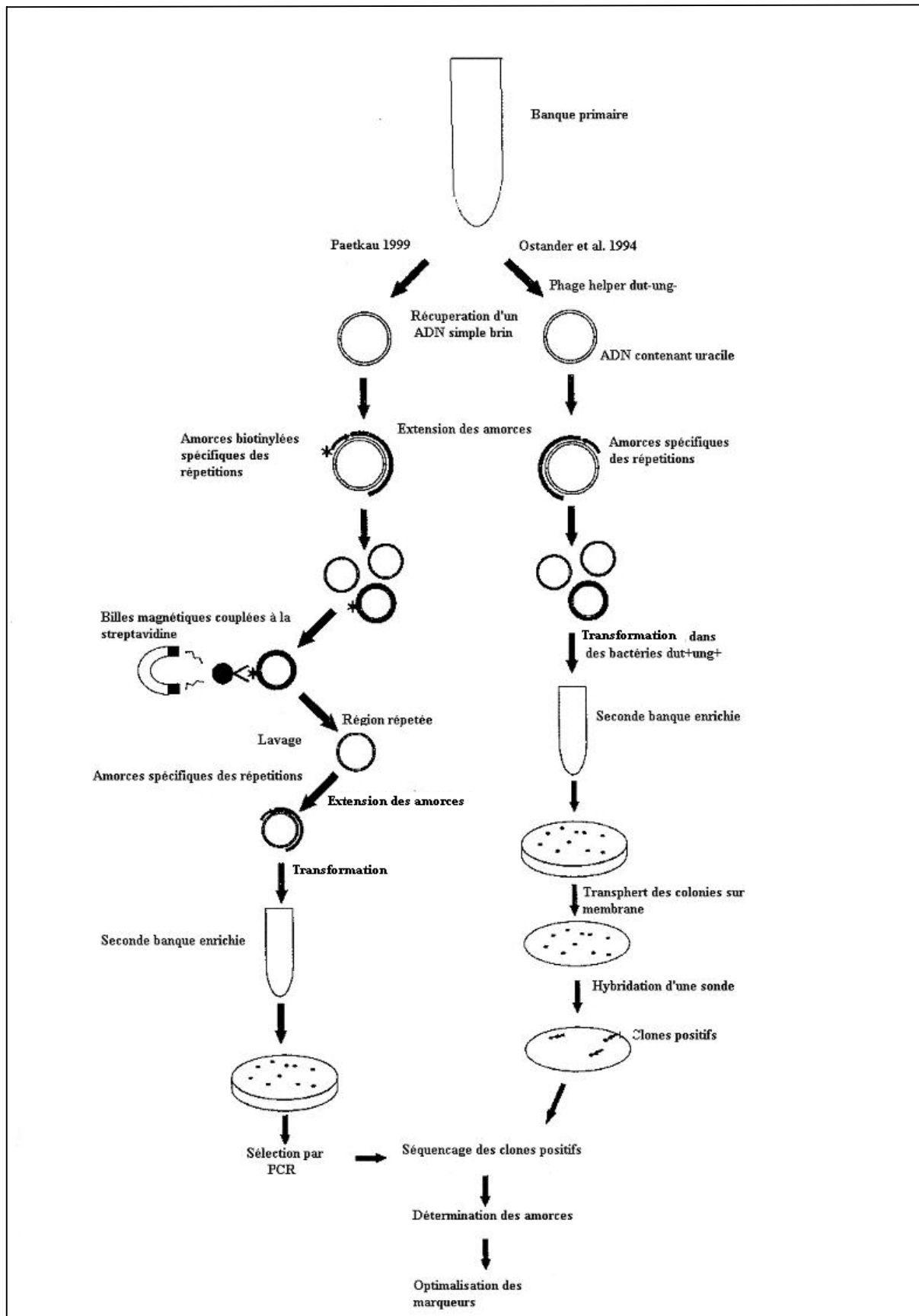


Fig. 15. Méthode d'isolation décrite par Olander et Paetkau. La méthode d'Olander utilise des amorces biotinylées spécifiques des régions répétées. Les fragments ainsi sélectionnés sont insérés dans une bactérie. Les bactéries positives sont criblées par une sonde. Dans la méthode décrite par Paetkau, l'ADN de la banque primaire est répliqué à l'aide de phage helper dans des bactéries *dut-ung-* dans le but d'obtenir de l'ADN simple brin contenant de l'Uracile. Seuls les fragments ayant amplifié à l'aide d'une sonde spécifique aux régions répétées pourront transformer les bactéries *dut+ung+*. Les bactéries contenant un microsatellite sont isolées à l'aide d'une sonde.

- L'ADN contenant les microsatellites est isolé grâce à une sonde qui est soit attachée à une membrane de nylon soit biotinylée et attachée à des billes magnétiques via la streptavidine.
- Après lavage et élution, les fragments d'ADN contenant les répétitions sont amplifiés par PCR, insérés dans un plasmide et utilisés pour la transformation des bactéries hôtes.
- Les clones positifs sont criblés soit par PCR, soit à l'aide d'une sonde, soit par séquençage.

La méthode que nous avons utilisée pour isoler des microsatellites est une application de l'hybridation sélective :

- Des adaptateurs ont été liés aux fragments d'ADN génomique.
- L'enrichissement s'est fait à l'aide de billes magnétiques couplées à la streptavidine ainsi qu'à l'aide d'une sonde biotinylée.
- Les clones positifs ont été isolés par hybridation de l'ADN sur membrane de nylon et par l'utilisation d'une sonde biotinylée s'hybridant de manière spécifique aux motifs microsatellites.

L'isolation de microsatellites dans le génome d'une espèce n'est qu'une première étape dans le développement de marqueurs microsatellites fonctionnels et polymorphes. En effet, il faut ensuite concevoir des amorces, optimiser les conditions d'amplification et trouver les loci microsatellites présentant du polymorphisme. Ainsi, au fur et à mesure, le nombre de marqueurs possibles diminue. Au final, seuls les marqueurs les plus polymorphes et les plus performants sont utilisés.

5. Objectifs

Le but de ce travail est d'isoler des microsatellites dans le génome de la chicorée industrielle et de définir les amorces qui pourront amplifier ces séquences. Les conditions d'amplification de ces amorces seront optimisées et seules les plus polymorphes seront sélectionnées. La dernière étape sera d'analyser la variabilité intrapopulation et interpolation pour six variétés de chicorée industrielle à l'aide des marqueurs microsatellites mis au point.

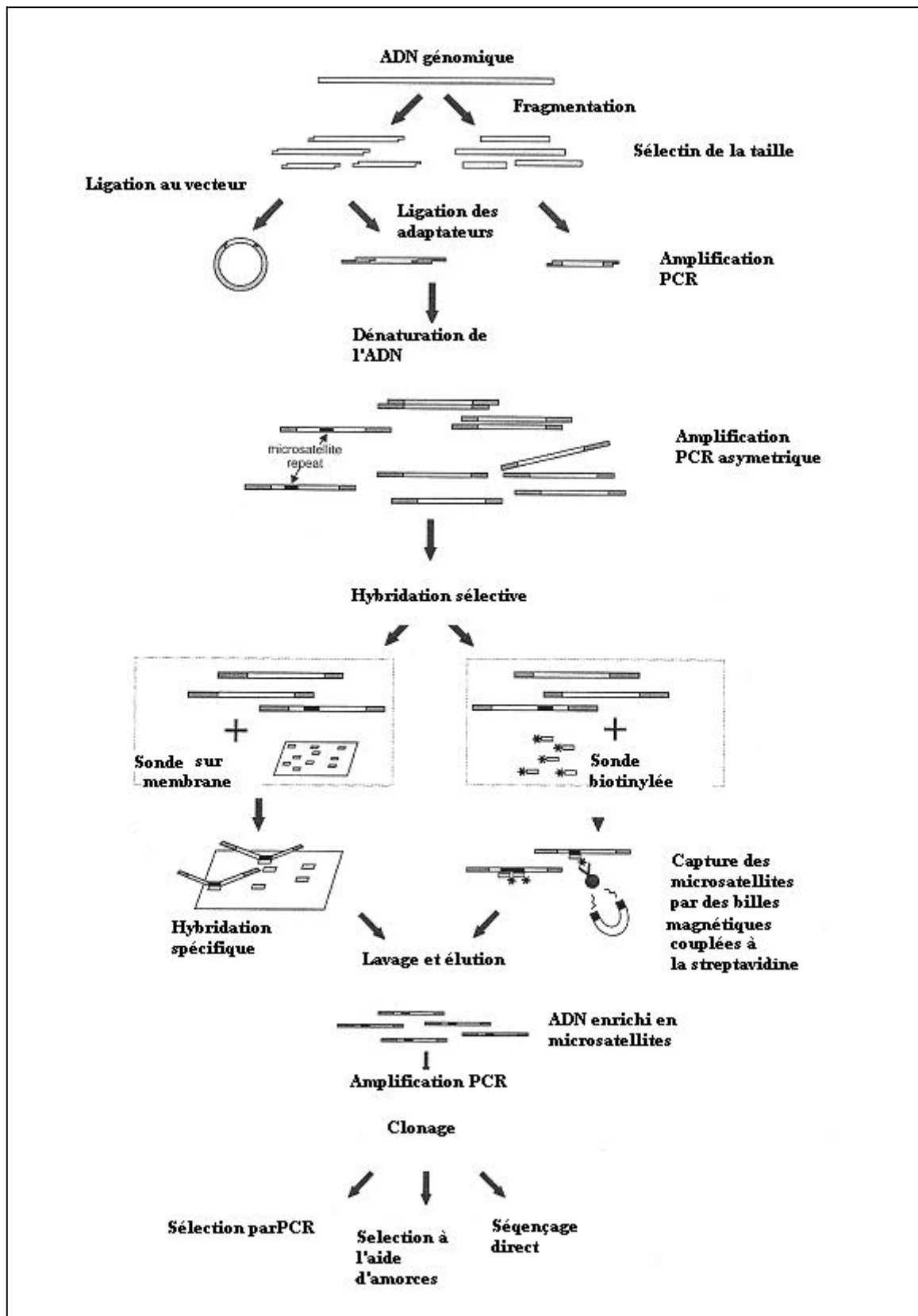


Fig. 16. Isolation de microsatellites par hybridation sélective. Les premières étapes sont identiques à la méthode classique. Les fragments d'ADN ainsi générés sont sélectionnés par une sonde soit hybridée au préalable sur une membrane soit biotynillée. L'ADN ainsi sélectionné est amplifié et cloné. Les bactéries positives sont sélectionnées soit par PCR soit à l'aide de sonde, soit par séquençage direct.

III. Matériel et méthodes

1. Extraction d'ADN génomique de chicorée

L'obtention de fragments nucléotidiques nécessite de l'ADN génomique de bonne qualité (sans contamination ni trace de dégradation). L'ADN est extrait à partir de racine de chicorée selon la méthode CTAB adaptée de Murray et Thompson (1981).

- Préchauffer à 60°C la solution de CTAB 2X :
Tris-HCl 100 mM, pH 8 (ACROS ORGANICS # 140500025), EDTA 20mM (SIGMA® # E-5134), NaCl 1,4 M (PROLABO #1723), CTAB (cethylmethyl ammonium bromide) 2% p/v (IGN Biomédical, Inc. #195029), PVP-40 soluble (polyvinylpyrrolidone) 1% p/v (SIGMA® # P-5288), 2-β mercaptoethanol 0,2% v/v (SIGMA® # M-6250), à ajouter uniquement au moment de l'utilisation.
- Mettre l'isopropanol au congélateur (-20°C).
- Refroidir le mortier et le pilon à l'azote liquide. Peser 3g de racine de chicorée et broyer l'échantillon en une fine poudre dans le mortier en présence d'azote liquide.
- Ajouter aux tissus broyés 22,5 ml de CTAB 2X et incubé sous agitation à 60°C pendant 1 heure. Mélanger de temps en temps et laisser refroidir l'échantillon à température ambiante.
- Extraire avec un volume de mélange Phénol:chloroforme:alcool isoamylique (25 :24 :1) (SIGMA® # P-2069) en agitant vigoureusement pendant 10 minutes à température ambiante. Centrifuger à 3300 rpm (2000 g) (X-22R Centrifuge, Beckman Coulter Allegra™) pendant 10 minutes à température ambiante et récupérer la phase aqueuse.
- Refaire l'étape d'extraction une deuxième fois.
- Précipiter en ajoutant 2/3 de volume d'isopropanol froid (-20°C). Homogénéiser, laisser 1 heure à température ambiante et centrifuger à 4000 rpm pendant 10 minutes à température ambiante.
- Eliminer le surnageant, ajouter 10 ml de tampon de lavage :
76% (v/v) ethanol, 10 mM ammonium acetate (SIGMA® # A-1542)
et laisser incubé 1 heure sous agitation à 4°C.
- Centrifuger à 2400 rpm (1000g) pendant 10 minutes à température ambiante puis éliminer le surnageant, sécher le culot (VR-1, Heto Lab Equipment) pendant environ 25 minutes et resuspendre dans 1 ml d'eau mili-Q stérile.
- Ajouter 2µl de RNase ADNse-free (SIGMA® # R-5503) et incubé 30 minutes à 37°C.
- Précipiter en ajoutant 2ml de TE (Tris 10 mM, pH 8 (ACROS ORGANICS # 140500025), EDTA 0,1 mM (SIGMA® # E-5134)), 1,1 ml d'acétate d'ammonium 7,5M (SIGMA® # A-1542) et 8,4 ml d'éthanol froid (96%). Homogénéiser par inversion et incubé à -70°C pendant 30 minutes.
- Centrifuger pendant 20 minutes à 3500 rpm, éliminer le surnageant et ajouter 1,5ml d'éthanol 70% froid. Réaliser cette étape 3 fois.
- Sécher le culot et suspendre dans 1ml d'eau mili-Q et stocker à -20°C.

Afin d'extraire de l'ADN génomique total de chicorée sur une quantité de départ de 100 mg de racine broyée, la quantité des solutions est divisée par trente hormis pour la RNase. Les différentes étapes sont réalisées dans des Eppendorf de 2 ml et les centrifugations se font dans la Centrifugeuse Biofuge 13 (Heraeus Sepatech)

2. Génération d'une collection de fragments d'ADN génomique

L'obtention d'une collection de fragments d'ADN génomique nécessite des segments d'ADN de taille comprise entre 400 et 900 pb. La fixation d'adaptateurs permettra de réaliser les étapes ultérieures.

2.1. Restriction de l'ADN génomique de chicorée

La restriction de l'ADN à l'aide de trois enzymes de restriction permet d'obtenir des fragments à partir desquels il sera possible d'isoler des séquences contenant des motifs microsatellites. Une sélection plus précise de la taille des fragments sera réalisée ultérieurement.

- Faire un mélange de 600 µl avec :

	Volume	Final
ADN génomique	200 µl	~200 µg
NEB2 10X	60µl	1X
<i>Hae</i> III 10U/µl (BioLabs #RO108S)	20µl	200U
<i>Rsa</i> I 10U/µl (BioLabs #RO167S)	20µl	200U
<i>Nhe</i> I 10U/µl (BioLabs #RO131S)	20µl	200U
H ₂ O		
Volume final	600µl	

- Vortexer et centrifuger légèrement (éviter d'avoir des gouttes sur les parois).
- Repartir 100 µl dans 6 tubes et incuber à 37°C pendant la nuit.

2.2. Traitement à la nucléase

Ce traitement permet d'obtenir des fragments à bouts francs.

- Rassembler les 6 tubes en un seul et inactiver les enzymes de restriction en incubant l'ADN à 65°C pendant 20 minutes.
- Ajouter ensuite directement à l'échantillon 5 µl de Mung bean Nucléase (50U) (BioLabs #M0250S) et laisser incuber pendant 30 minutes à 30°C.
- Purifier les fragments d'ADN à l'aide du kit « microcon, Centrifugal Filter Devices » (Millipore #42413) conformément aux instructions du fabricant.

2.3. Traitement à la phosphatase alcaline

Les fragments d'ADN sont ensuite déphosphorylés à la phosphatase alcaline (CIAP).

- Faire le mélange suivant :

	Volume	Final
Fragments d'ADN génomique	125µl	
Tampon CIAP 10x (NEB2)	30 µl	1x
H ₂ O	140µl	
Phosphatase (BioLabs # MO290S)	5 µl	50U
Volume total	300 µl	

- Incuber l'ADN pendant 2h00 à 37°C.
- Purifier les fragments d'ADN à l'aide du kit « microcon, Centrifugal Filter Devices » (Millipore #42413) conformément aux instructions du fabricant.

2.4. Récupération des fragments d'ADN entre 400 et 900 pb avec le kit QiaExII

- Charger 2 puits d'un gel d'agarose 1% avec 40 µl d'ADN digéré + 10 µl de bleu de chargement. Ajouter en parallèle avec un puits d'écart, un marqueur d'échelle.
- Après migration, découper sous UV la région comprise entre 400 et 900 pb et mettre les morceaux de gel dans un Eppendorf.
- Récupérer les fragments à l'aide du kit QiaExII (QIAGEN #20021) suivant les instructions du fabricant.

2.5. Hybridation des amorces SNX pour former l'adaptateur double brin

Amorces SNX reverse : 5'-GCT-TCT-GCT-AGC-AAG-GCC-TTA-GAA-AA-3'

Amorces SNX forward : 5'-CTA-AGG-CCT-TGC-TAG-CAG-AAG-C-3'

- Dans un tube, mélanger 100 µl de chaque amorce (SNX forward et SNX reverse, 100 µM) de manière à avoir une concentration finale en adaptateur double brin de 50 µM.
- Placer le tube contenant les deux amorces dans un Erlen de 1l contenant de l'eau en ébullition et laisser incuber pendant 10 minutes à cette température.
- Laisser refroidir lentement jusqu'à température ambiante.

2.6. Réaction de ligation/restriction simultanée de l'adaptateur SNX aux fragments génomique

➤ Réaction de ligation/restriction.

	Volume	Final
ADN génomique	40µl	
Linker SNX double brin(50 µM)	6 µl	5 µM
Tampon T4 ligase (10x stock)	12 µl	1x
H ₂ O	54,8 µl	
BSA (100x stock)	1,2 µl	1x
Enzyme de restriction <i>Xmn</i> I (20U/µl) (BioLabs #R0194S)	4 µl	80U
T4 DNA Ligase (high concentration : 2000U/µl) (BioLabs #M0202S)	2 µl	4000U
Volume final:	120 µl	

La ligation se fait durant la nuit dans un appareil PCR avec le profil de température suivant : 30 cycles définis comme suit : 30 minutes à 16°C et 10 minutes à 37°C. La ligation se termine par une inactivation des enzymes par un passage de 20 minutes à 65°C.

➤ Vérification de la ligation des adaptateurs par PCR

Si la PCR avec les amorces SNX donne un produit d'amplification, cela signifie que les adaptateurs sont bien liés aux fragments d'ADN génomique.

	Volume	Final
Produit de ligation	2 µl	
H ₂ O	36,5 µl	
Tampon PCR 10x	5 µl	1x
DNTP mixture (10 mM de chaque nucléotide) (Roche #1581295)	1 µl	0,2mM
SNX forward 10 µM	5 µl	1 µM
Taq polymerase (5U/ µl) (BioLabs #MO267L)	0,5 µl	2,5U/µl
Volume total :	50 µl	

Profil de température : 5 minutes à 95°C, ensuite 30 cycles définis comme suit : 30 secondes à 95°C, 1 minute à 55°C, 1 minute à 72°C. On termine l'amplification par une élongation de 7 minutes à 72°C.

3. Enrichissement des fragments génomiques en séquences contenant des motifs microsatellites

3.1. Hybridation soustractive

L'étape d'enrichissement sera réalisée à l'aide de billes magnétiques couplées à la streptavidine, reconnaissant une sonde biotynilée GA₁₅ spécifique des séquences microsatellites GA_n.

- Fixation de l'oligo GA₁₅-biotine à la streptavidine se trouvant à la surface des billes magnétiques
- Préparer une solution de SSC 20x (NaCl 3M, Citrate de Na 0,3 M (FLUKA #71404), pH 7). A partir de cette solution, faire les dilutions suivantes : SSC 5x, SSC 2x, SSC 2x 0,1% SDS (Sigma # L-5750), SSC 1x, SSC 1x 0,1% SDS.
 - Resuspendre les billes et en prélever 100 µl dans un Eppendorf.
 - Les billes sont lavées à trois reprises avec 150 µl de SSC 5x
 - Retirer le SSC 5x et faire le mélange suivant :

	Volume	Final
{GA} ₁₅ -biotine 100 µM	6 µl	3 µM
Streptavidine magnetic particles (Kit PolyA Tract mRNA Isolation System II # 25300, Promega)	1 mg	1 mg
SSC 20x	50 µl	5x
H ₂ O	144 µl	
Volume total :	200 µl	

- Incuber 15 minutes à température ambiante sous agitation.
- Enlever l'excès d'oligomères non liés par 3 lavages successifs avec 200 µl de milieu SSC 5x et resuspendre les billes dans 70 µl de SSC 20x.

- Hybridation de l'oligo {GA}₁₅-biotine aux fragments d'ADN
 - Après ligation des adaptateurs, diluer 20 µl d'ADN_{gen} (de taille comprise entre 400 et 900 pb) dans 110 µl d'eau, dénaturer pendant 10 minutes à 95°C et mettre directement sur glace.
 - Ajouter à l'ADN dénaturé les 70 µl de billes magnétiques couplées à la sonde GA₁₅.
 - L'hybridation est réalisée pendant 45 minutes à 55°C.
 - Laver : 2x avec 200µl de SSC 2x à température ambiante pendant 5 minutes, 4x avec 200µl de SSC 2x, 0.1% SDS à 62°C pendant 5 minutes, 4x avec 200µl de SSC 1x, 0.1% SDS à 62°C pendant 5 minutes, 2x avec 200µl de SSC 1x à 62°C pendant 5 minutes.
 - Eluer les fragments d'ADN génomique contenant des séquences microsatellites de type {GA}_n en ajoutant 40 µl d'une solution de NaOH (20 µl de NaOH 1N, 5 µl de NaCl 4M, 175 µl d'eau) et laisser incubé 20 minutes à température ambiante sous agitation.
 - Ajouter ensuite 20 µl d'HCl 0,2M et 4,4 µl de TE 10x (Tris 100 mM (pH 8) (ACROS ORGANICS # 140500025), EDTA 1 mM (SIGMA® # E-5134)).
 - L'échantillon est ensuite dessalé sur une colonne microcon (Millipore #42413) et élué avec 20 µl d'eau.
- Enrichissement par PCR des fragments d'ADN génomique purifiés par hybridation avec un oligonucléotide biotinylé

	Volume	Final
Produit d'éluion	5 µl	
H ₂ O	33,5	
Tampon PCR 10x	5µl	1x
DNTP mixture (10 mM de chaque nucléotide)	1 µl	0,2 mM
SNX forward 10 µM	5 µl	1 µM
Taq polymérase (5U/µl)	0,5 µl	2,5 U
Volume total :	50 µl	

Profil de température : 5 minutes à 95°C, ensuite 40 cycles définis comme suit : 30 secondes à 95°C, 1 minute à 55°C, 1 minute à 72°C. On termine l'amplification par une élongation de 7 minutes à 72°C.

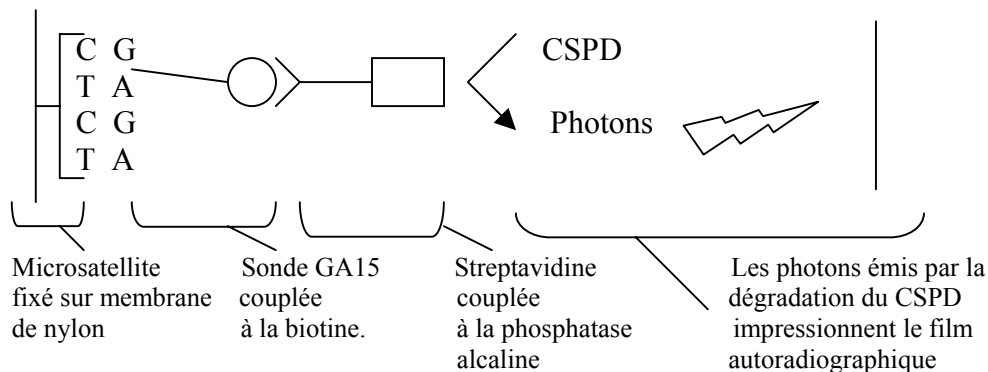
3.2. Insertion des fragments d'ADN_{gen} enrichis en microsatellites dans le Vecteur pCR® II et transformation dans TOP10F'

- Insertion des fragments dans le vecteur PCR II
Le vecteur PCR II est un vecteur de type TA. Cela signifie que le T en 3' du vecteur s'hybride avec le A ajouté par la Taq polymérase lors de l'amplification des fragments d'ADN génomique. Le fragment s'insère dans le gène LacZ.
- Insertion des plasmides dans une bactérie TOP10F'
Les bactéries possédant un fragment d'ADN génomique dans leur plasmide sont détectées par leur couleur blanche sur un milieu contenant du X-gal (en effet, le gène LacZ est inactif)

Voir les instruction du fabricant. (TA Cloning® Kit Dual Promoteur, Invitrogen #K02060-40).

3.3. Criblage sur membrane de nylon

Cette étape vise à isoler les clones contenant effectivement des séquences microsatellites. Pratiquement, l'ADN des colonies transformées (blanches) est transféré sur une membrane de nylon. L'emplacement des clones contenant une séquence microsatellite sera détecté grâce à la fixation d'une sonde GA₁₅ biotynilée au microsatellite GA_n. Cette sonde se fixera à son tour à la streptavidine elle-même couplée à la phosphatase alcaline. La phosphatase alcaline dégrade un substrat chémiluminescent (CSPD), ce qui provoque l'émission de photons qui impressionnent un film autoradiographique.



- Repiquage des clones sur boîte de Pétri
 - Placer une boîte de Pétri contenant du milieu LB additionné d'ampicilline (éviter les contaminants extérieurs) sur une matrice quadrillée avec 50 ou 100 emplacements.
 - Repiquer les clones blancs et incuber environ 16 heures à 37°C.
- Transfert de l'ADN des clones sur membrane
 - Les membranes (HybondTM-N⁺, Amersham #RPN203B) doivent être manipulées avec des gants et des pinces.
 - Avant le transfert, les membranes de nylon sont numérotées conformément aux numéros des boîtes et orientées à l'aide d'un système à trois points (traits disposés asymétriquement au stylo bic)
 - Déposer chaque membrane sur la boîte de Pétri (marquage sur la face supérieure) pendant 5-10 minutes et reporter très précisément les marques de la membrane sur le côté de la boîte de Pétri (Pas sur le couvercle !).
 - Pendant ce temps, imbiber jusqu'à saturation 3-4 feuilles de papier Whatman 3M avec la solution de lyse des bactéries et de dénaturation de l'ADN (NaOH 0,5 M, NaCl 1,5 M).
 - Enlever délicatement à l'aide d'une pince souple les membranes des boîtes de Pétri, et les transférer sur le Whatman imbibé de solution (face avec les bactéries à l'air).
 - Laisser les membranes pendant 10 minutes.
 - Pendant ce temps, imbiber jusqu'à saturation 3-4 feuilles de papier Whatman 3M avec la solution de neutralisation (Tris HCl 0,5 M pH 7,5, NaCl 1,5 M).
 - Transférer les membranes à l'aide d'une pince plate sur ces feuilles (face avec les bactéries à l'air).
 - Laisser les membranes pendant 10 minutes.
 - Pendant ce temps, imbiber jusqu'à saturation 3-4 feuilles de papier Whatman 3M avec la solution de SSC2x (NaCl 300 mM, citrate de Na 30 mM, pH 7).
 - Transférer les membranes à l'aide d'une pince plate sur ces feuilles (face avec les bactéries à l'air).
 - Laisser les membranes 5 minutes.

- Récupérer les membranes sur des feuilles de papier Whatman sec (face avec les bactéries à l'air).
 - Incuber au four pendant 1 heure à 80°C pour fixer l'ADN.
 - Les membranes ainsi transférées peuvent être conservées à température ambiante pendant très longtemps (il est cependant nécessaire de les protéger de la lumière).
- Criblage par hybridation sur membrane
- Préparer les solutions suivantes :
 - ❖ SDS 10% (10 ml) : 1 ml de SDS, 9 ml d'eau
 - ❖ SSC 20x (0,5l) : 50,25 gr de trisodium citrate déshydraté (FLUKA #71404), 87,65 gr de NaCl
 - ❖ Solution de rinçage R1 (1L) : 100 ml de SSC 20x (SSC 2x), 1 ml de SDS (0,1%), 899 ml d'eau
 - ❖ Solution de rinçage R2 (1L) : 50 ml de SSC 20x (SSC 1x), 1 ml de SDS (0,1%), 949 ml d'eau
 - ❖ Tampon de lavage (1l) : 13,81 gr d'acide maléique (0,1 M) (Fulka #63190), 8,76 gr de NaCl (0,15 M), ajuster à pH 7,5, 3 ml de Tween 20 (0,3% v/v)(Fulka #93773)
 - ❖ Tampon T2 (1l) : même chose que tampon T1 sans Tween 20
 - ❖ Tampon de détection (1l) : Tris HCl 0,1 M, NaCl 0,1, Ajuster à pH 9,5
 - ❖ Solution stock de blocage : A préparer la veille
Dissoudre 10 gr de Blocking reagent (produit du kit « Biotin Detection » : Roche – # 1 811 592) dans 100 ml de tampon T2, chauffer au micro-ondes et autoclaver.
Ouvrir stérilement
 - ❖ Tampon d'hybridation (100 ml) : 25 ml de SSC 20x, 10 ml de solution stock de blocage, 200 µl de SDS 10%, 100 mg de N- Lauroylsarkosine (Sigma® #L-9150), 65 ml d'eau
 - ❖ Solution de blocage 5x (80 ml) : **A préparer frais**, 40 ml de solution stock de blocage, 40 ml de tampon T2
 - ❖ Solution de streptavidine-AP : A préparer frais, 20 ml de solution de blocage 5x, 4 µl de streptavidine-AP (produit du kit Biotin Detection : Roche – reference : 1 811 592). Centrifuger pendant 10 minutes à 4000 rpm avant chaque usage et prélever la quantité nécessaire.
 - Placer les membranes (face bactérienne à l'air) dans un tube à hybridation (2 membranes par tube).
 - ***Etape de déprotéination***
50 ml par tube de solution R1 avec 500 µl de protéinase K (Sigma® #P-2308) à 20 mg/ml. Incubation 50 minutes à 45°C.
 - ***Rinçages*** : 2x 10 minutes à 45°C dans 50 ml de solution R1.
 - ***Préhybridation*** : 2 heures à température d'hybridation (55°C) dans 50 ml de tampon d'hybridation.
 - ***Hybridation*** : 10 à 16 heures (une nuit) à 55°C, dans 10 ml de tampon d'hybridation dans lesquels on a rajouté 1 µl de mix de sonde marquée à la biotine.
 - ***Rinçages*** : 2x 12 minutes à 55°C dans 50 ml de solution R1.
2x 10 minutes à 55°C dans 50 ml de solution R2.
5 minutes à 55°C dans 50 ml de tampon de lavage T1.
 - ***Blocage des sites aspécifiques*** : 30 minutes à température ambiante dans 35 ml de solution de blocage
 - ***Incubation dans la solution de streptavidine-AP*** : Incubation 30 minutes à température ambiante dans 20 ml de solution streptavidine-AP

- **Rinçages** : 2 fois 15 min. à température ambiante dans 100 ml de tampon de lavage T1
5 minutes à température ambiante dans 20 ml de tampon de détection
- **Etape de détection**
Mettre les membranes sur un film plastique et le recouvrir (avec environ 2 ml) de CSPD ready-to-use (produit du kit Biotin Detection : Roche – #1 811 592). Refermer le film plastique sur la membrane.
Laisser incuber pendant 5 minutes à température ambiante.
Eliminer l'excès de liquide et placer les membranes sur un deuxième film en enfermant la membrane.
Placer les membranes dans l'étuve à 37°C.
Exposer les membranes pendant 2 minutes sur un film radiographique (Kodak #1435726) sensible à la lumière.
- **Repiquage des clones positifs.**

4. Séquençage des clones positifs

4.1. Miniprep

- Repiquer les bactéries transformées sur milieu solide et laisser à 37°C pendant 24h.
- Préparer les solutions suivantes :

<u>Tampon P1 (stocké à 4°C)</u>		
Rnase A	100 µg/ml	20 mg/200 ml
Tris HCl	50 mM pH 8,0	1,2114 gr/200 ml
EDTA	10 mM	0,744 gr/200 ml (SIGMA® # E-5134).
<u>Tampon P2 (stocké à température ambiante)</u>		
NaOH	200 mM	1,6 gr/200 ml
SDS	1 % (P/V)	2 gr/200 ml
<u>Tampon P3 (stocké à 4°C)</u>		
Kac	3,0 M pH 5,5	58,89 gr/200 ml. (Sigma #P-5708)
- Prendre une plaque multipuits destinée aux cultures liquides de bactéries et ajouter dans chaque puits 1,2 ml de milieu « CIRCLE GROW » (Q-BIO gene #3000-112) contenant de l'ampicilline (100 ng/ml)
- Mettre dans chaque puits un clone en phase de croissance exponentielle repiquée la veille.
- Mettre à 37°C sous agitation pendant 1 nuit.
- Réaliser un stock glycérol des transformants à partir de la culture liquide :
 1. mettre 120 µl de glycérol 40 % (filtré à 0,22µm) dans chaque puits d'une plaque de stockage 96 puits.
 2. ajouter dans chaque puits 120 µl de culture liquide.
 3. stocker à -70°C.
- Centrifuger (Refrigered Centrifuge PK120R, ALC) le reste de la culture liquide pendant 10 minutes à 3500 rpm.
- Retirer le surnageant, ajouter 300 µl de tampon de lyse (tampon P1) et homogénéiser.
- Laisser incuber 10 minutes, ajouter 300 µl de tampon P2 (tampon de lyse) et homogénéiser.
- Laisser incuber 10 minutes et ajouter 300 µl de tampon P3 (précipitation des protéines).

- Laisser incubé 30 minutes et centrifuger pendant une heure à 3500 rpm.
- Prendre une nouvelle plaque multipuits, et placer au dessus d'elle une plaque multiscreen (Millipore #MAGVN2210) servant de filtre ne laissant pas passer les protéines.
- Mettre 200 µl de la première plaque dans les filtres multiscreen et centrifuger 3 minutes à 1000 rpm. Refaire 2 fois la dernière opération.
- Ajouter 800 µl d'isopropanol dans la deuxième plaque multipuits et laisser incubé 30 minutes à température ambiante.
- Centrifuger 1h30 à 3500 rpm, enlever l'isopropanol, laver avec 500 µl d'éthanol 70%, centrifuger et laisser sécher une nuit à 37°C afin que tout l'éthanol s'évapore.
- Ajouter dans chaque puits 25 µl d'eau, centrifuger 30 secondes, laisser 1h00 sous agitation et stocker dans une plaque multipuits – 20°C

4.2. Séquençage

Matériel :

- CEQ DTCS quick start kit (Beckman #608120)
- CEQ Sequencing reaction buffer kit (Beckman #608074)
- Stop solution : 1 µl de glycogène (Roche #901303), 2 µl de d'acétate de sodium 3 M (sigma #S-7899), 1,6 µl d'eau, 0,4 EDTA 0,5 M (sigma # E-7889)
- CEQ Sample loading solution (SLS) (Beckman #608082)
- SEQ sequencing separation buffer (Beckman #608012)
- Séquenceur Beckman CEQ 8000

Procédure :

- Préparer un mélange contenant 1 µl d'amorce (10 pmoles /µl), X µl d'ADN (à déterminer suivant la quantité d'ADN). Ajuster le volume à 10 µl avec de l'eau stérile sans DEPC.
- Dénaturer le mélange à 95 °C pendant 15 minutes puis mettre immédiatement sur glace pendant 3 minutes.
- Rajouter 4 µl de CEQ DTCS quik start kit, 1 µl de CEQ Sequencing reaction buffer kit, 5 µl d'eau milliQ.
- Lancer une réaction de séquençage avec le programme suivant : 92°C pendant 1'30'' ; 40 cycles définis comme suit : 20'' à 96°C ; 20'' à 50°C ; 4' à 60°C. Ensuite, 4°C à l'infini.
- Purifier en ajoutant au 20 µl de solution de séquençage : 5 µl de Stop solution, 60 µl d'éthanol 96 % froid.
- Centrifuger 15 minutes à 13 000 g puis éliminer le surnageant.
- Laver avec 200 µl d'éthanol 76 %, centrifuger 15 minutes à 13 000 g et éliminer le surnageant, refaire cette étape une seconde fois.
- Sécher au speedvac pendant 25 minutes.
- Resuspendre dans 40 µl de SLS. Placer dans un puits d'une plaque de séquençage et couvrir avec une goutte d'huile minérale.
- Prendre une nouvelle plaque, et ajouter dans chaque puits 8 gouttes de SEQ sequencing separation buffer. Placer cette plaque et celle contenant les échantillons à l'endroit prévu dans l'appareil de séquençage.
- Lancer la méthode de séparation LFR-a (CEQ 8000 Genetic analysis system) dont les paramètres sont repris ci-dessous :
 - température des capillaires lors de la migration : 50°C
 - température de dénaturation : 90°C durée : 120 secondes
 - injection : voltage : 2 kV, durée : 15 secondes
 - séparation voltage : 4 kV, durée : 110 minutes
- Analyser les séquences avec les paramètres d'analyse par défaut.

5. Design des amorces

- Intégrer la séquence dans le programme « Primer 3 » (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) et sélectionner les paramètres suivants :
 - primer size : 18<20<27
 - primer Tm : 57<60<63
 - primer GC% : 40<50<60
 - number to return : 10
- Calculer le ΔG pour la formation d'hairpins ainsi que de duplex pour chaque amorce à l'aide du logiciel « Oligo » et sélectionner les amorces ayant le ΔG le moins négatif.

6. Optimisation des conditions d'amplification pour chaque paire d'amorces

- Pour chaque amorce, réaliser 8 tubes PCR contenant chacun :

	Volume	Final
ADN _{gén.}	1 μ l	
H ₂ O	40,6	
Tampon PCR 10x	5 μ l	1x
DNTP mixture (10 mM de chaque nucléotide)	1 μ l	0,2 mM
Amorce FW	1 μ l	1 μ M
Amorce RV	1 μ l	
Taq polymérase (5U/ μ l)	0,5 μ l	2,5 U
Volume total :	50 μ l	

- Programmer les étapes suivantes dans l'appareil PCR (iCycler thermal cycler, Biorad) :
95°C pendant 4 minutes, ensuite 35 cycles définis comme suit : 95°C pendant 30 secondes, de 55°C à 65°C pendant 30 secondes (gradient de température), 75°C pendant 1 minute. Terminer l'amplification par une élongation à 75°C pendant 7 minutes.
Le gradient de température est réparti sur huit puits de la manière suivante :
Puits 1 : 65°, Puits 2 : 64,3°C, Puits 3 : 63,1°C, Puits 4 : 61,2°C, Puits 5 : 58,7°C, Puits 6 : 57°C, Puits 7 : 55,8°C, Puits 8 : 55°C.
- Faire migrer les huit produits PCR sur un gel d'acrylamide 12% :
 - **solution de séparation** (pour 1 gel) : 1,6 ml d'eau, 2 ml d'acrylamide 30% (BioRad # 161-0158), 1,3 ml de tris pH 8,8, 0,05 ml d'APS 10% (Chimica #2015374), 0,002 ml de TEMED (Sigma #T-8133).
 - **solution d'empilement** : (pour 1 gel) 2,7 ml d'eau, 670 μ l d'acrylamide 30%, 500 μ l de Tris pH 8,8, 40 μ l d'APS 10%, 10 μ l de TEMED.
- Déterminer la température optimale d'hybridation pour l'amorce, c'est-à-dire la température pour laquelle l'amplification est maximale.

7. Etude du polymorphisme

- Pour chaque amorce, réaliser une PCR sur 6 ADN différents. Pour chaque PCR faire la solution suivante :

	Volume	Final
ADN _{gén.}	0,5 µl	
H ₂ O	20,3 µl	
Tampon PCR 10x	2,5 µl	1x
DNTP mixture (10 mM de chaque nucléotide)	0,5 µl	0,2 mM
Amorce FW	0,5 µl	1 µM
Amorce RV	0,5 µl	
Taq polymerase (5U/µl)	0,2 µl	2,5 U
Volume total :	25 µl	

- Faire migrer les 6 produits PCR sur un gel d'acrylamide 12% et sélectionner les amorces ayant le plus de variabilité pour la bande correspondant à la taille attendue.
- Pour chaque paire sélectionnée, recommander des amorces couplées à un fluorochrome WELLRED (Beckman) distribué par Proligo (www.proligo.com).

8. Détermination de la taille des allèles microsatellites

- Réaliser une PCR pour chaque amorce fluorescente et pour chaque ADN purifié.
Mélange PCR :

	Volume	Final
ADN _{gén.}	0,5 µl	
H ₂ O	20,3 µl	
Tampon PCR 10x	2,5 µl	1x
DNTP mixture (10 mM de chaque nucléotide)	0,5 µl	0,2 mM
Amorce FW	0,5 µl	1 µM
Amorce RV	0,5 µl	
Taq polymerase (5U/µl)	0,2 µl	2,5 U
Volume total :	25 µl	

- Dans une plaque de séquençage, ajouter 40 µl de SLS (CEQ Sample loader solution, Beckman #608082), 0,3 µl de standard de taille (CEQ Beckman Coulter #608090), 2 à 4 µl de produit PCR et couvrir avec 1 goutte d'huile minérale (CEQ Beckman Coulter #608114).
- Dans une autre plaque, ajouter dans chaque puits correspondant, 8 gouttes de tampon de séparation (SEQ sequencing separation buffer, Beckman #608012).
- Placer les deux plaques dans le séquenceur (Séquenceur Beckman CEQ 8000) et démarrer la migration.
- Après analyse, répertorier la taille et le nombre des différents allèles.

IV. Résultats et discussion

1. Clonage de fragments d'ADN génomique enrichis en microsatellites

1.1. Isolement de l'ADN génomique

L'obtention d'une collection de fragments d'ADN génomique (lignée 97TM17-03) en vue d'isoler des microsatellites requiert une grande qualité de matériel de départ. Les données ci-dessous rendent compte de la qualité et de la quantité de l'ADN obtenu (*Matériel et méthode 1.*).

Analyse au spectrophotomètre : Estimation de la concentration : 1,32 µg/µl
Pureté (rapport 260/280 nm) : 1,9

Migration sur gel d'agarose :

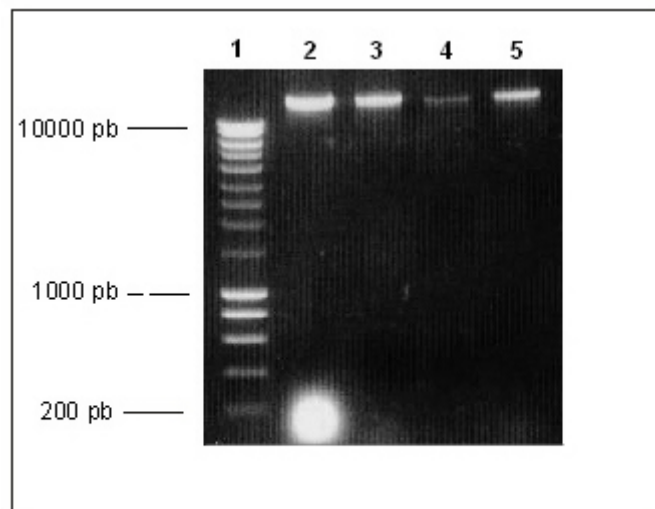


Fig.1. Analyse sur gel d'agarose 1% de l'ADN génomique obtenu. 5 µl de Smart Ladder (1). 5 µl d'ADN de Chicorée purifié mais non traité à la RNase (2). ADN purifié après traitement à la RNase. 5 µl (3), 1 µl (4), 3 µl (5).

La différence entre les puits 2 et 3 montre que le traitement à la RNase a bien fonctionné. La quantité totale d'ADN génomique isolé est estimée à 1320 µg.

1.2. Traitement de l'ADN génomique en vue de lier les adaptateurs SNX aux fragments d'ADN

L'ADN purifié doit subir plusieurs traitements afin de permettre la ligation des adaptateurs SNX.

- Restriction par les enzymes de restriction HaeIII, RsaI et NheI (*Matériel et méthode 2.1.*)
Cette étape permet de fragmenter l'ADN génomique en segments de petite taille (entre 200 et 1000 pb)
- Traitement à la nucléase (*Matériel et méthode 2.2.*)

Les deux enzymes de restriction HaeIII et RsaI génèrent directement des bouts francs. Par contre l'enzyme NheI produit des bouts protrudants. La nucléase enlève les extensions simple brin générées par cette enzyme. Pour la ligation de l'adaptateur, tous les fragments doivent posséder des bouts francs.

- Traitement à la phosphatase alcaline (*Matériel et méthode 2.3.*)
Ce traitement permet de retirer les groupements phosphates à l'extrémité 5' des fragments. Cette étape facilitera la ligation des adaptateurs car elle empêche les fragments de se lier entre eux.
- Sélection des fragments de taille entre 400 et 900 pb (*Matériel et méthode 2.4*)
Les fragments trop grands ne permettent pas de séquencer la totalité du fragment en une fois et, à l'opposé, les fragments de trop petite taille laisseront trop peu de liberté pour la sélection d'amorces flanquantes.

Les figures 2a et 2b montrent l'étape de sélection des fragments d'ADN compris entre 400 et 900 paires de bases.

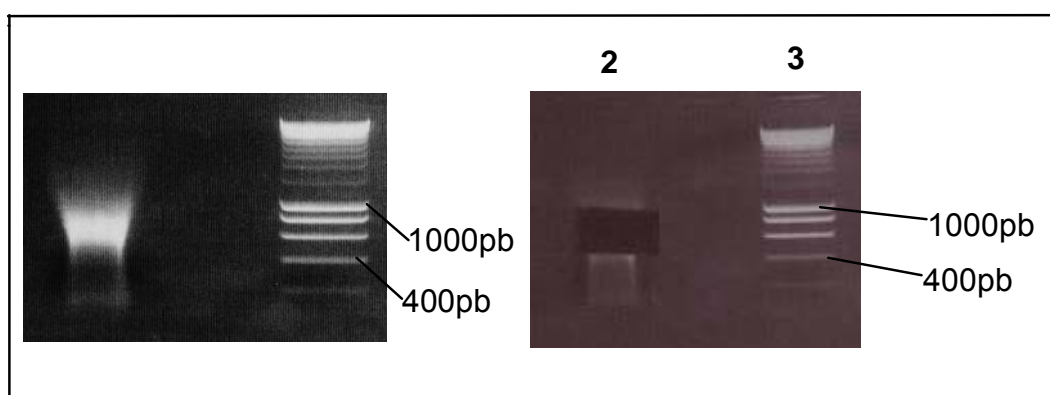


Fig. 2. Migration sur gel d'agarose 1% de l'ADN génomique après restriction avec les enzymes HaeIII, RsaI, NheI, traitement à la nucléase et à la phosphatase alcaline. **(1)** 40 µl d'ADN génomique traité. **(2)** 5 µl de Smart Ladder. **(3)** Visualisation de l'excision de la bande contenant les fragments entre 400 et 900 pb.

La taille des fragments générés est, déjà avant sélection, majoritairement entre 400 et 800 pb. Cette étape est cependant importante car elle permet d'éliminer tous les petits fragments qui pourraient prendre le dessus lors des amplifications.

Les fragments sont donc prêts à être liés à l'adaptateur (*Matériel et méthode 2.6.*). L'ajout de ces adaptateurs a pour but de permettre l'amplification après l'hybridation soustractive afin d'augmenter le nombre de fragments contenant un microsatellite.

Pour favoriser cette ligation, l'enzyme XmnI est ajoutée dans le milieu. En effet, la formation d'un dimer d'adaptateurs SNX crée un site de restriction pour l'enzyme XmnI. La présence de XmnI permet donc de séparer les adaptateurs et dès lors de favoriser la ligation entre l'adaptateur et le fragment d'ADN et d'éviter la formation de dimères d'adaptateurs.

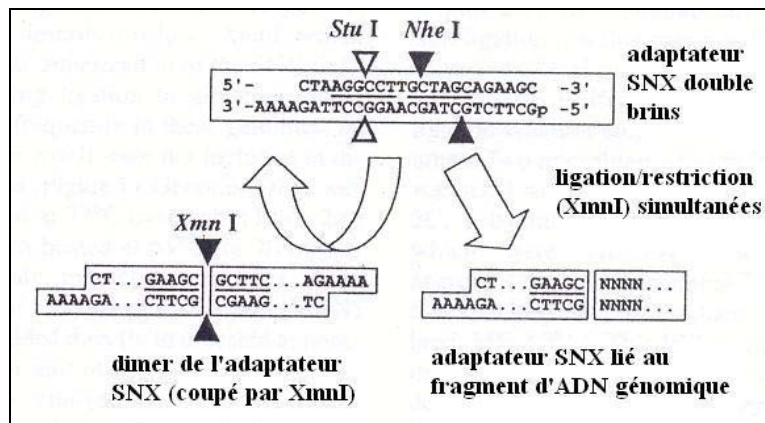


Fig. 3. Principe de la ligation des adaptateurs SNX.

Un contrôle de ligation a été réalisé en procédant à une amplification PCR utilisant des amorces spécifiques des adaptateurs. Le produit d'amplification a montré, lors de son passage sur gel, une traînée, prouvant que la ligation a bien eu lieu.

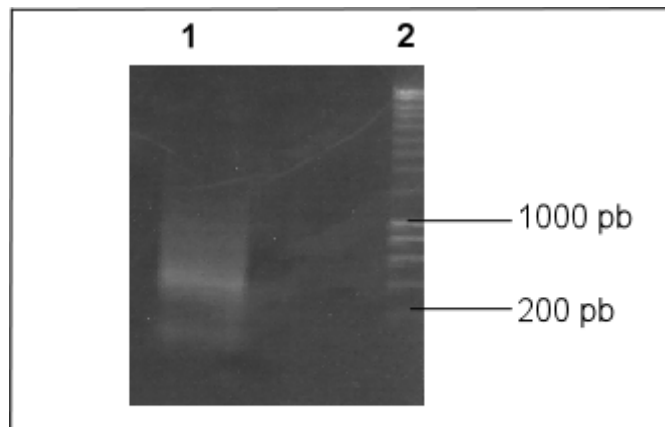


Fig. 4. Migration sur gel d'agarose 1% du produit PCR après ligation des adaptateurs SNX aux fragments génomiques de chicorée. **(1)** 50 µl de produit d'amplification avec les amorces spécifiques des adaptateurs. **(2)** 5 µl de Smart Ladder.

1.3. Enrichissement de la collection de fragments d'ADN en séquences microsatellites

L'étape déterminante pour l'isolement de microsatellites est l'hybridation soustractive (*Matériel et méthode 3.1.*).

Les fragments d'ADN génomique contenant une séquence microsatellite sont isolés grâce à l'hybridation avec une sonde GA₁₅ biotinylé. Cette sonde est au préalable couplée avec des billes magnétiques associées à la streptavidine. Un aimant permettra de récupérer les billes magnétiques ainsi que les fragments d'ADN associés à celles-ci.

La collection de fragments d'ADN enrichie en microsatellite ainsi obtenue subit les traitements suivants :

- Amplification par PCR :
Grâce aux adaptateurs fixés aux fragments d'ADN, les simples brins sélectionnés par hybridation soustractive sont amplifiés par des amorces spécifiques des adaptateurs SNX. Cette étape permet d'obtenir des fragments doubles brins et d'augmenter la taille de notre collection en séquences microsatellites.
- Insertion des fragments dans le vecteur PCRII (Matériel et méthode 3.2).
Le vecteur PCRII est un vecteur de type TA. Cela signifie que le T en 3' du vecteur s'hybride avec le A ajouté par la Taq polymérase lors de l'amplification des fragments d'ADN génomique. Le fragment s'insère dans le gène LacZ.

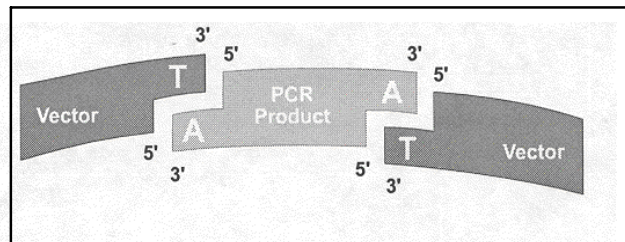


Fig. 5. Principe de l'insertion d'un fragment d'ADN dans un vecteur TA.

- Insertion des plasmides dans une bactérie TOP10F'
Les bactéries possédant un fragment d'ADN génomique dans leur plasmide sont détectées par leur couleur blanche sur un milieu contenant du X-gal (le gène LacZ est rendu inactif par la présence de l'insert)
- Criblage des clones contenant un fragment microsatellite dans leur vecteur (Matériel et méthode 3.3).
L'ADN des colonies transformées (blanches) est transféré sur une membrane de nylon. L'emplacement des clones contenant une séquence microsatellite sera détecté grâce à la fixation d'une sonde GA₁₅ biotinylée au microsatellite GA_n. Cette sonde se fixera à son tour à la streptavidine qui est associée à la phosphatase alcaline. La phosphatase alcaline dégrade un substrat chimiluminescent (CSPD), ce qui provoque l'émission de photons impressionnant un film autoradiographique (Fig.6.)

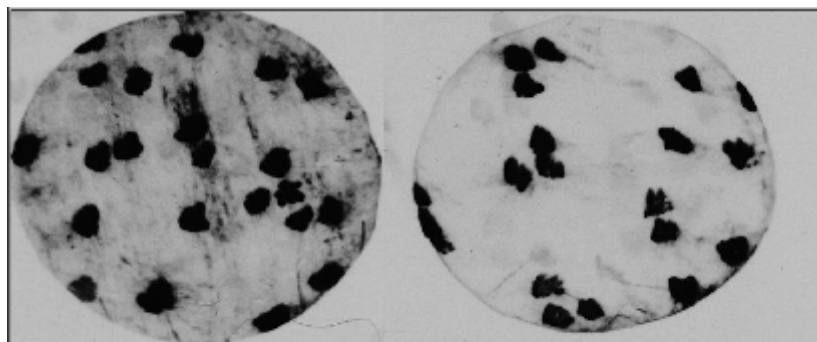


Fig. 6. Films autoradiographiques montrant l'emplacement des clones contenant dans leur vecteur une séquence microsatellite.

Ce n'est qu'à la fin du criblage sur membrane que l'on sait si l'enrichissement a fonctionné. Dans le cas contraire, toutes les étapes doivent être refaites à partir de l'hybridation soustractive.

Au total, le protocole de l'hybridation soustractive a été modifié trois fois. Les modifications ont été réalisées sur deux paramètres influençant la spécificité de l'enrichissement :

- La température : La spécificité de l'hybridation sera d'autant plus grande que la température est élevée. Cependant, à partir de la température de melting (T_m) l'hybridation ne peut plus avoir lieu. La température intervient lors de l'hybridation et des lavages.
- La concentration en sel : La spécificité de l'hybridation est d'autant plus grande que la concentration en sel est faible. Cette concentration est représentée par la concentration en SSC (SSC 1x : 150 mM NaCl, 15 mM citrate de sodium)

	Hybridation	Lavages	Clones positifs	%
A	30 min. à 55°C	- 4x 200µl SSC 2x, 0.1% SDS à t° ambiante. - 4x 200µl SSC 1x, 0.1% SDS à 65°C - 2x 200µl SSC 1x à 65°C	4/100 (test 1) 2/100 (test 2)	4% 2%
B	1 nuit à 65°C	- 4x 500µl SSC 2x, 0.1% SDS à t° ambiante. - 4x 500µl SSC 1x, 0.1% SDS à 65°C - 1x 500µl SSC 1x à 65°C - 1x 500µl SSC 0.2x à 65°C	0/100 (test 1)	0%
C	45 min. à 55°C	- 2x SSC 2x à t° ambiante - 4x 200µl SSC 2x, 0.1% SDS à 62°C - 4x 200µl SSC 1x, 0.1% SDS à 62°C - 2x 200µl SSC 1x à 62°C	12/150 (test 1) 56/350 (test 2)	8% 16%

Tableau 1. Modifications apportées au protocole d'enrichissement en fragments microsatellites lors des étapes d'hybridation et de lavage. Le nombre de clones positifs lors du criblage sur membrane de nylon ainsi que le pourcentage correspondant sont indiqués pour chacun des protocoles (A, B, C).

Dans le protocole A, l'hybridation se fait à une température de 55°C, la concentration en SSC est diminuée jusqu'à 1x et les lavages se font à 65°C. Le pourcentage maximum de clones contenant une séquence microsatellite après le criblage est de 4%. Ce pourcentage signifie qu'il y a encore 96 fragments sur 100 qui ne possèdent pas de microsatellite. La spécificité du protocole d'enrichissement n'est pas suffisante.

Afin de tenter de remédier à ce problème, pour le protocole B, la température et le temps d'hybridation ont été augmentée et la concentration en sel diminuée (jusqu'à SSC 0,2%). La stringence de l'hybridation est donc plus élevée que pour le premier protocole. Aucun clone positif n'a été détecté. Il semblerait donc que les conditions d'hybridation et/ou de lavage soient trop stringentes pour permettre aux fragments de s'attacher à la sonde et d'y rester pendant les lavages.

Pour le protocole C, la température d'hybridation est de 55°C, la concentration en SSC ne descend pas en dessous de 1X et la température des lavages est de 62°C. Suite à ce protocole, le pourcentage de clones positifs lors du criblage (8 à 16%) est jusqu'à quatre fois supérieur au premier protocole. Les conditions mises en place ont donc permis d'améliorer l'enrichissement de la collection de fragments d'ADN en séquences microsatellites.

Au final, sur les 800 clones criblés sur membrane, 76 ont été visualisés sur film autoradiographique.

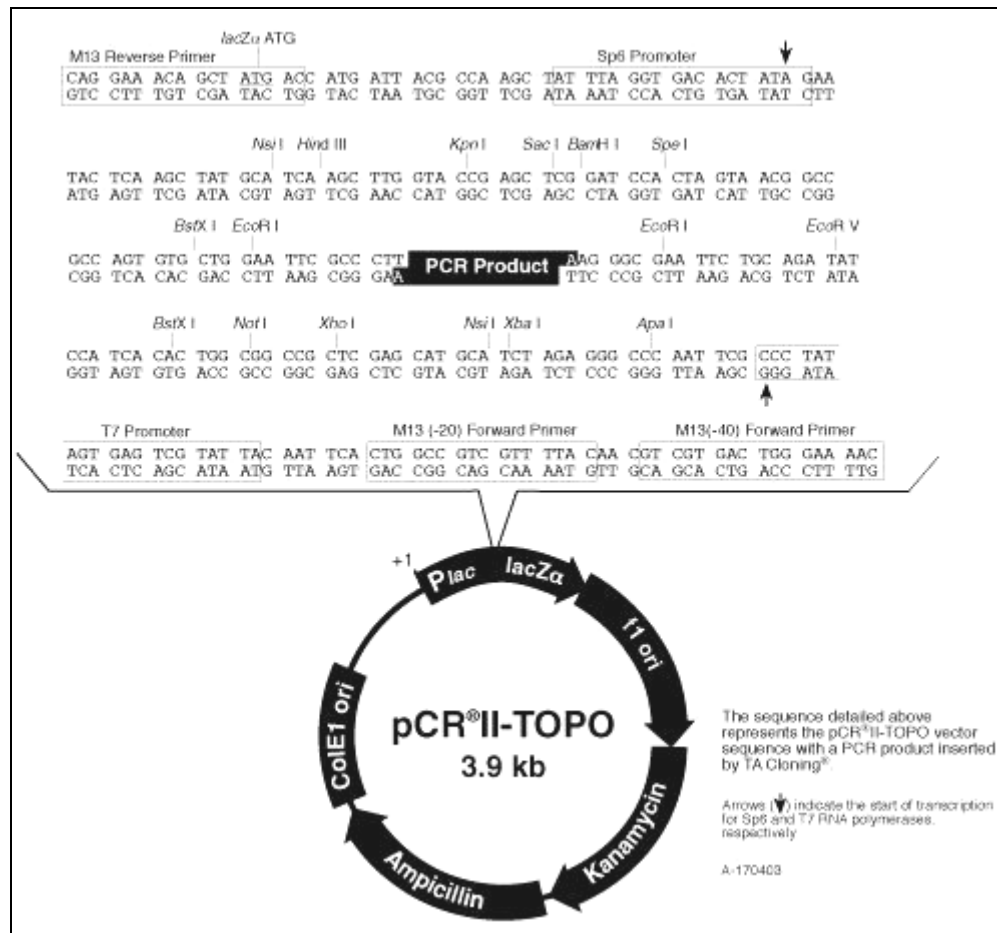


Fig. 7. Séquence du plasmide PCR[®]II

2. Détermination des amorces

2.1. Sélection des clones à séquencer

Afin d'extraire le plasmide de la bactérie, une minipréparation (*Matériel et méthode 4.1.*) a été effectuée sur les 76 clones positifs lors du criblage. Ensuite, chaque plasmide a subi une restriction par EcoRI permettant ainsi de libérer l'insert (Fig. 7.) et d'estimer sa taille sur gel d'agarose.

Le tableau 2 présente l'estimation de la taille de l'insert pour les 76 clones.

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
Insert (pb)	420	170	400	500	ND	ND	380	ND	300	260	200	100	280	250	ND	ND
Séquencé	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Insert (pb)	300	400	300	ND	260	550	320	400	500	220	300	200	330	200	380	470
Séquencé	X	X	X			X	X	X	X		X		X		X	X

	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Insert (pb)	ND	500	250	400	240	350	230	ND	500	270	250	210	220	220	ND	280
Séquencé		X		X	X	X	X		X	X						

	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
Insert (pb)	580	350	200	200	410	ND	350	310	200	ND	320	ND	100	330	360	310
Séquencé	X	X			X		X	X	X		X			X	X	X

	65	66	67	68	69	70	71	72	73	74	75	76
Insert (pb)	100	50	300	300	270	300	380	300	350	500	250	350
Séquencé			X	X	X	X	X	X	X	X		X

Tableau 2. Estimation de la taille de l'insert par digestion avec l'enzyme EcoRI pour les 76 clones et recensement des clones séquencés.

La taille moyenne estimée des inserts est de 310 pb.

Les 19 premiers clones ont été séquencés directement avant que la taille de l'insert ne soit estimée par restriction du plasmide. Pour les clones suivants, les inserts séquencés ont été sélectionnés sur base de leur taille respective et de leur concentration lors de la miniprep.

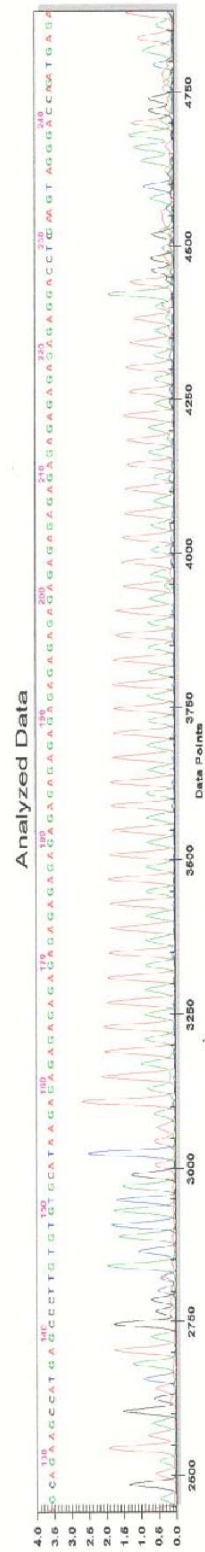


Fig. 8. Séquençage d'un microsatellite GA33.

2.2. Séquençage des clones sélectionnés

53 clones ont été sélectionnés pour le séquençage (*Matériel et méthode 4.2*).

Le tableau 3 détaille, pour chaque clone, le brin séquencé, la taille réelle de l'insert (déterminée après séquençage), ses caractéristiques et le nombre de répétitions.

Clone	Fw	Rv	Taille	Caractéristiques	GA _n	Clone	Fw	Rv	Taille	caractéristiques	GA _n
01	X	X	338	Pas de microsat.		34	X	X	428	Utilisable	29
02	X		ND	Pas de microsat		36	X	X	339	Utilisable	44
03	X	X	372	Bord trop court	32	37	X		ND	Bord trop court	32
04	X	X	442	Utilisable	12	38	X	X	309	Utilisable =76=62	30
05	X		ND	Bord trop court	10	39	X		ND	Bord trop court	32
06	X	X	221	Utilisable	36	41	X	X	469	Utilisable	12
07	X	X	477	Utilisable	23	42	X		ND	Pas analysé	27
09	X	X	249	Bord trop court		49	X	X	ND	Bord trop court	32
10	X	X	238	Bord trop court	37	50	X	X	ND	Pas d'assemblage	43
11	X	X	150	Utilisable	32	53	X	X	337	Utilisable	24
12	X	X	ND	Bord trop court	22	55	X		ND	Pas analysé	27
13	X	X	ND	Pas d'assemblage	15	56	X	X	255	= 64	21
14	X	X	ND	Pas d'assemblage	41	57	X		ND	Pas analysé	33
15	X	X	321	Utilisable	35	59	X		ND	Pas analysé	15
16	X	X	ND	Pas d'assemblage	17	62	X	X	309	=76=3 8	30
17	X	X	ND	Pas d'assemblage	46	63	X	X	ND	Pas d'assemblage	43
18	X	X	359	Utilisable	11	64	X	X	255	Utilisable	21
19	X	X	ND	Pas d'assemblage	14	67	X	X	ND	Bord trop court	44
22	X	X	478	Utilisable	40	68	X		ND	Pas de microsat	
23	X	X	283	Utilisable	17	69	X	X	ND	Bord trop court	36
24	X	X	357	Utilisable	21	70	X		ND	Pas analysé	24
25	X	X	447	Utilisable	33	71	X		ND	Pas analysé	28
27	X		ND	Bord trop court	10	72	X	X	207	Utilisable	24
29	X		ND	Bord trop court	45	73	X		ND	=50	43
31	X		ND	Pas microsat.	44	74	X		ND	Pas de microsat	
32	X	X	402	Utilisable	32	76	X		309	=62=38	30

Tableau 3. Détail de l'analyse du séquençage des clones positifs. **Fw** : le brin forward a été séquencé. **Rv** : le brin reverse a été séquencé. **Taille** : taille de l'insert en paires de base. **Caractéristiques** : *Bord trop court* : les régions bordant le microsatellite ne sont pas suffisamment grandes pour déterminer des amorces. *Pas d'assemblage* : les séquences forward et reverse ne s'assemblent pas, probablement à cause d'un problème de séquençage. *Pas de microsat* : la séquence ne comporte pas de microsatellite. *Pas analysé* : la séquence reverse n'a pas été séquencée pour des raisons de temps et donc l'assemblage n'a pas pu être réalisé. *Utilisable* : les séquences forward et reverse ont été assemblées correctement et les régions bordantes sont suffisamment grandes pour déterminer des amorces. Cette séquence peut donc être utilisée afin de déterminer des marqueurs microsatellites. **GA_n** : nombre de répétition GA que comporte le microsatellite.

Taille moyenne des inserts	340 pb	
Inserts sans microsatellite	6/53	11,3%
Inserts avec régions bordantes trop courtes	12/53	22,6%
Séquences non assemblées	7/53	13,2%
Séquences non analysées	6/53	11,3%
Séquences non utilisables	35/53	66%
Séquences utilisables	18/53	34%
Nombre moyen de répétitions	29	

Tableau 4. Synthèse des informations figurant dans le tableau 3.

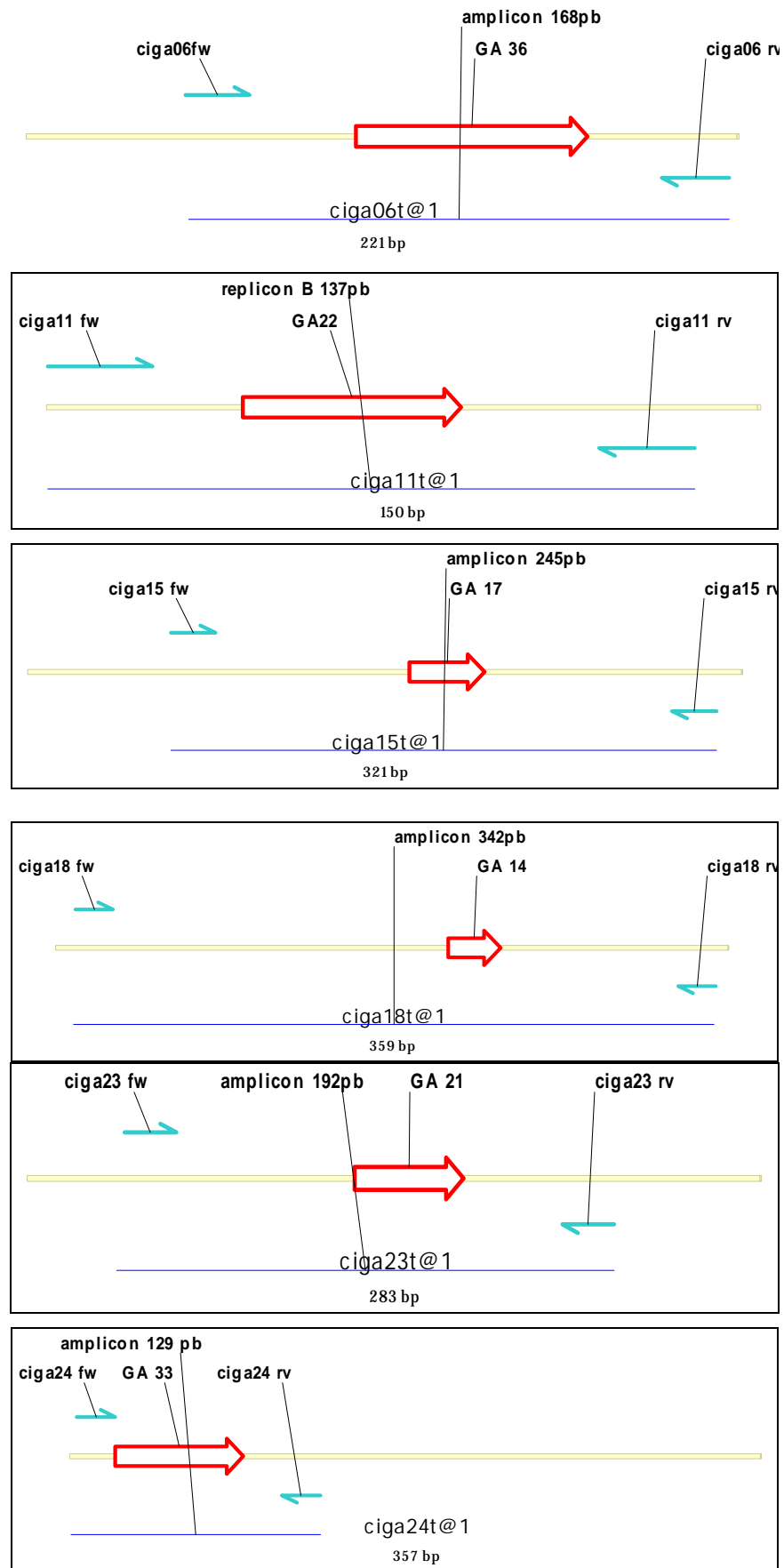


Fig. 9.a. Représentation schématique des séquences microsatellites avec la position des amorces Forward et Revers, la taille, la position du microsatellite et la taille de l'amplicon.

2.3. Design des amorces

Les paires d'amorces ont été déterminées sur les 18 séquences utilisables reprises dans le tableau 4 (*Matériel et méthode 5.*). Le tableau ci-dessous détaille, pour chaque amorce :

- le nom de la séquence à partir de laquelle les amorces ont été déterminées,
- le ΔG le plus contraignant : le programme informatique « Oligo » calcule pour chaque paire d'amorces la valeur énergétique lors de la formation de duplex entre les amorces et d'épingles à cheveux au sein d'une amorce. Plus cette valeur est négative, plus la conformation est stable. Parmi tous les ΔG calculés pour une amorce, seule la valeur la plus négative est représentée,
- le nombre de répétitions GA dans le microsatellite,
- la taille attendue lors de l'amplification,
- si la paire d'amorces a été sélectionnée.

Nom de la séquence	ΔG le plus contraignant	Nombre de répétitions GA	Taille de l'amplicon	Amorces sélectionnées
ciga04	-4,4	12	105	
ciga15	-5,1	17	245	X
ciga34	-5,3	29	160	X
ciga64	-6,3	21	217	X
ciga72	-6,3	24	189	X
ciga06	-6,5	36	168	X
ciga18	-6,8	14	342	X
ciga41	-6,8	12	169	
ciga24	-6,9	33	129	X
ciga53	-7,5	24	249	X
ciga23	-7,8	21	192	X
ciga25	-7,9	10	233	
ciga32	-8,2	32	301	X
ciga38	-8,3	39	177	X
ciga11	-8,4	22	137	X
ciga07	-9,9	23	290	
ciga22	-10,9	28	199	
ciga36	-10,9	44	210	

Tableau 5. Informations relatives aux paires d'amorces déterminées à partir du programme Primer 3.

La sélection des amorces se base sur deux critères :

- **Eviter la formation de dimères et d'épingles à cheveux stables :** ce sont donc les paires d'amorces ayant le ΔG le moins négatif qui ont été choisies.
- **Favoriser les microsatellites avec un grand nombre de répétitions :** le taux de mutation est proportionnel au nombre de répétitions du microsatellite. Donc, plus le microsatellite est grand, plus la probabilité que celui-ci soit polymorphe est élevée.

Au total, 12 paires d'amorces ont été sélectionnées afin de tester leur polymorphisme. La figure 9 est une représentation schématique des séquences microsatellites avec la position des amorces Forward et Reverse, la taille, la position du microsatellite et la taille de l'amplicon.

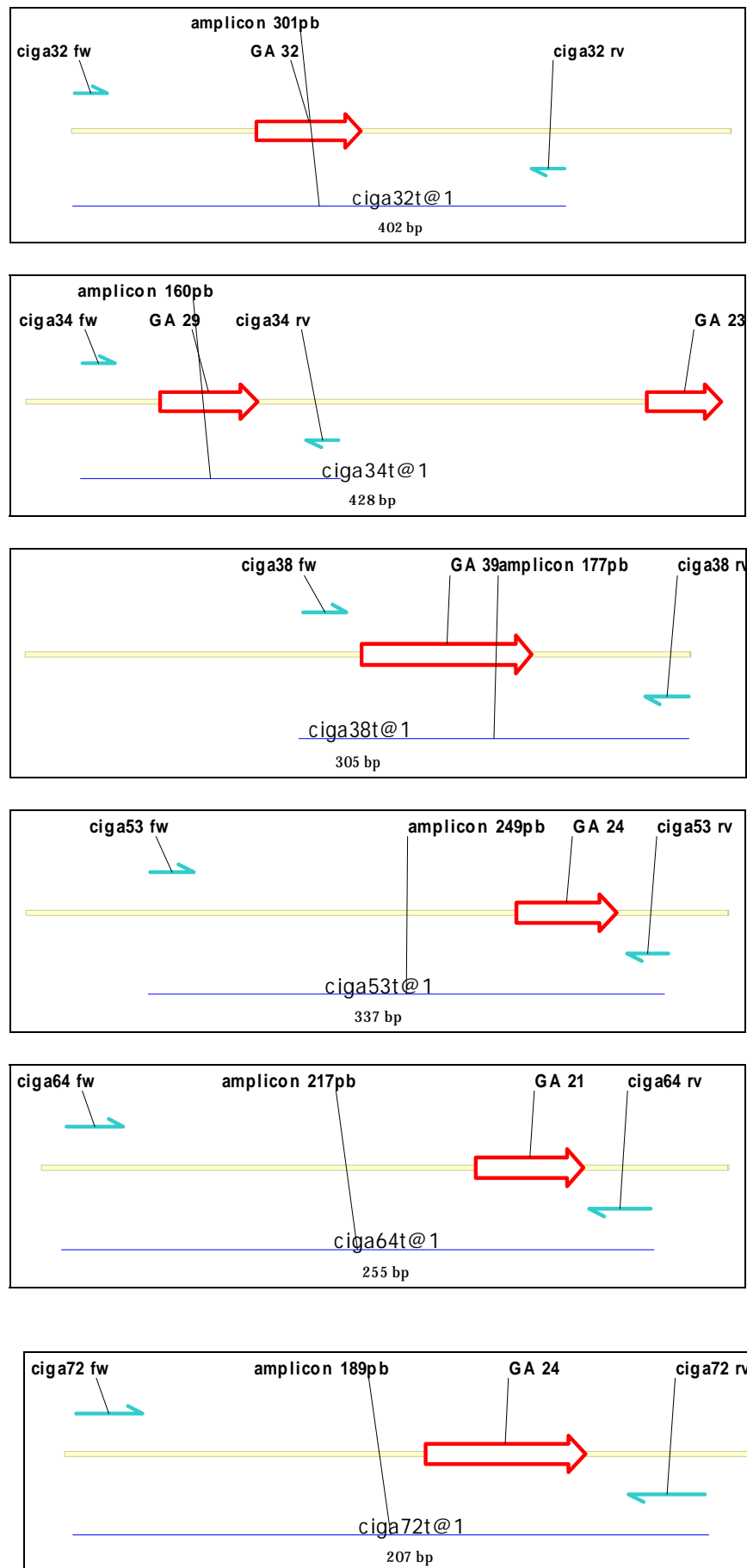


Fig. 9. b.Représentation schématique des séquences microsatellites avec la position des amorces Forward et Revers, taille, la position du microsatellite et la taille de l'amplicon.

2.4. Test d'amplification

La température optimale d'hybridation des amorces lors de la PCR a été déterminée à l'aide d'un gradient de température (*Matériel et méthode 6.1*).

Le tableau ci-dessous indique, pour chaque paire d'amorces, à quelle température d'hybridation une amplification a été observée.

	65°C	64,3°C	63,1°C	61,2°C	58,7°C	57°C	55,8°C	55°C
Amorces 6	X	X	X	X	X	X	X	X
Amorces 11			X	X	X	X	X	X
Amorces 15	X	X	X	X	X	X	X	X
Amorces 18	X	X	X	X	X	X	X	X
Amorces 23	X		X	X		X	X	X
Amorces 24			X	X	X	X	X	X
Amorces 32					X		X	
Amorces 34	X		X		X	X		
Amorces 38								
Amorces 53	X	X			X	X		
Amorces 64	X	X	X	X	X	X	X	X
Amorces 72	X	X	X	X	X	X		

Tableau 6. Températures d'hybridation de chaque paire d'amorces pour lesquelles ont observé une amplification .

L'amplification a lieu majoritairement à une température d'hybridation de 57 et 58,7°C. Lors de la détermination des amorces à l'aide du programme primer3, un des paramètres était l'obtention d'amorces ayant un T_m aux alentours de 60°C. Afin de minimiser les amplifications non spécifiques, c'est la température la plus élevée qui a été choisie, soit 58,7°C.

La figure 10 est un exemple de migration sur gel d'acrylamide 12% des 8 produits PCR de la paire d'amorces 18 pour des températures d'hybridation allant de 65 à 55°C. La coloration a été réalisée avec du bromure d'ethidium. La bande attendue se situe aux alentours de 350 pb. Les autres bandes sont des amplifications non spécifiques. L'augmentation de la température ainsi que la dénaturation des produits PCR avant chargement sur gel n'ont eu aucun effet sur ces bandes parasites.



Fig. 10. Migration sur gel d'acrylamide 12% de 8 produits d'amplification de la paire d'amorces 18 sur l'ADN de chicorée (variété: Nausica) pour des températures d'hybridation allant de 55 à 65°C.

2.5. Etude du polymorphisme

Afin de sélectionner les amorces les plus polymorphes, l'ADN génomique de six variétés de chicorée industrielle a été amplifié avec les 12 paires d'amorces sélectionnées. Les 72 produits obtenus ont ensuite migré sur un gel d'acrylamide 12%. La révélation s'est faite en plongeant les gels dans un bain de bromure d'éthidium (*matériel et méthode 6.*).

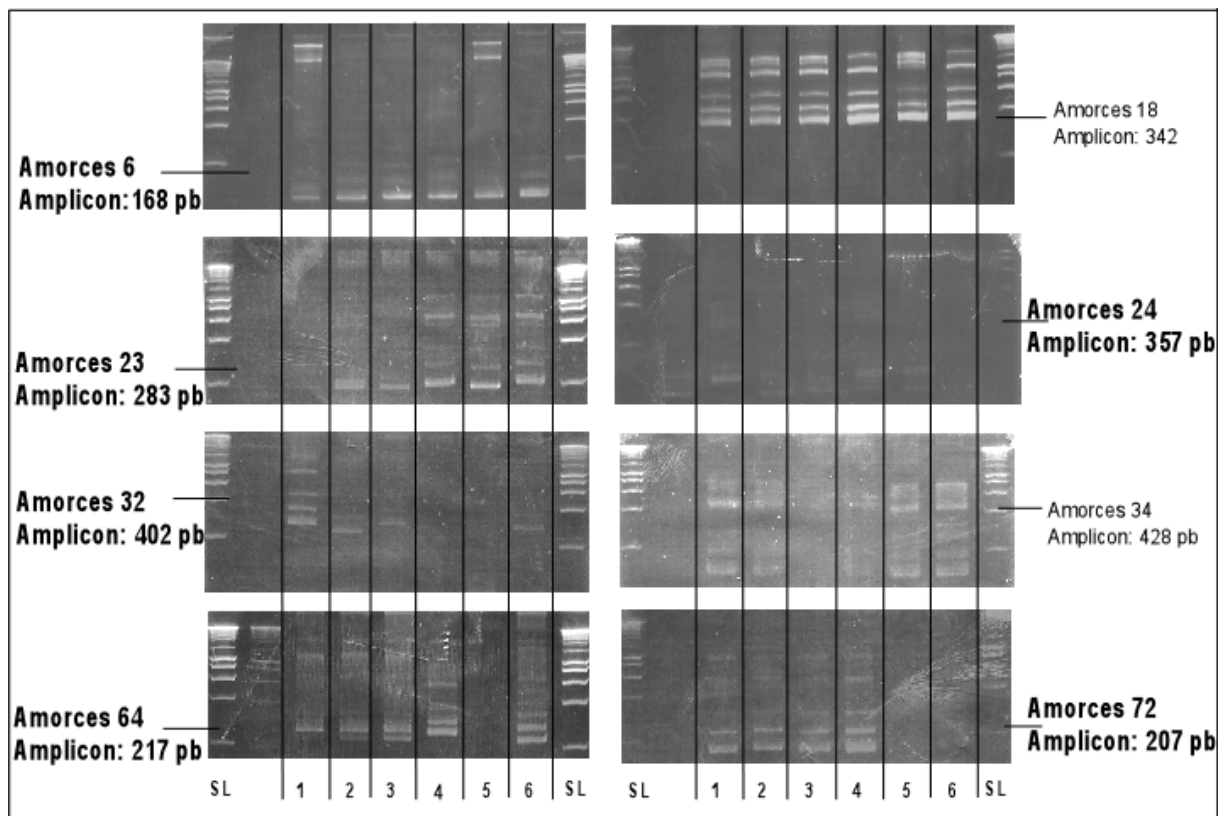


Fig. 11. Migration sur gel d'acrylamide 12% des produits PCR réalisés pour chaque paire d'amorces et pour 6 ADN différents. SL: Smart Ladder. 10 µl de produit PCR avec l'ADN génomique de la variété Nausica (1), Melci (2), Arancha (3), Madonna (4), Vivace (5) et la lignée 97TM17-03 (6).

Quatre paires d'amorces n'ont donné aucun produit d'amplification. Elles ont donc été écartées. Pour les 8 autres paires d'amorces, la bande de taille attendue a été comparée entre les 6 variétés de chicorée industrielle afin de déterminer le polymorphisme du marqueur étudié (Fig.11.)

Les produits d'amplification des amorces 6, 23, 64 et 72 montrent une variation de taille entre les variétés pour la bande attendue.

Les bandes de taille attendue pour les produits d'amplification des amorces 32 et 24 sont peu visibles. Par contre, il semblerait que leur profil général de migration diffère d'une variété à l'autre.

Les produits d'amplification des amorces 18 et 34 semblent moins polymorphes que les produits d'amplification des 6 autres paires d'amorces.

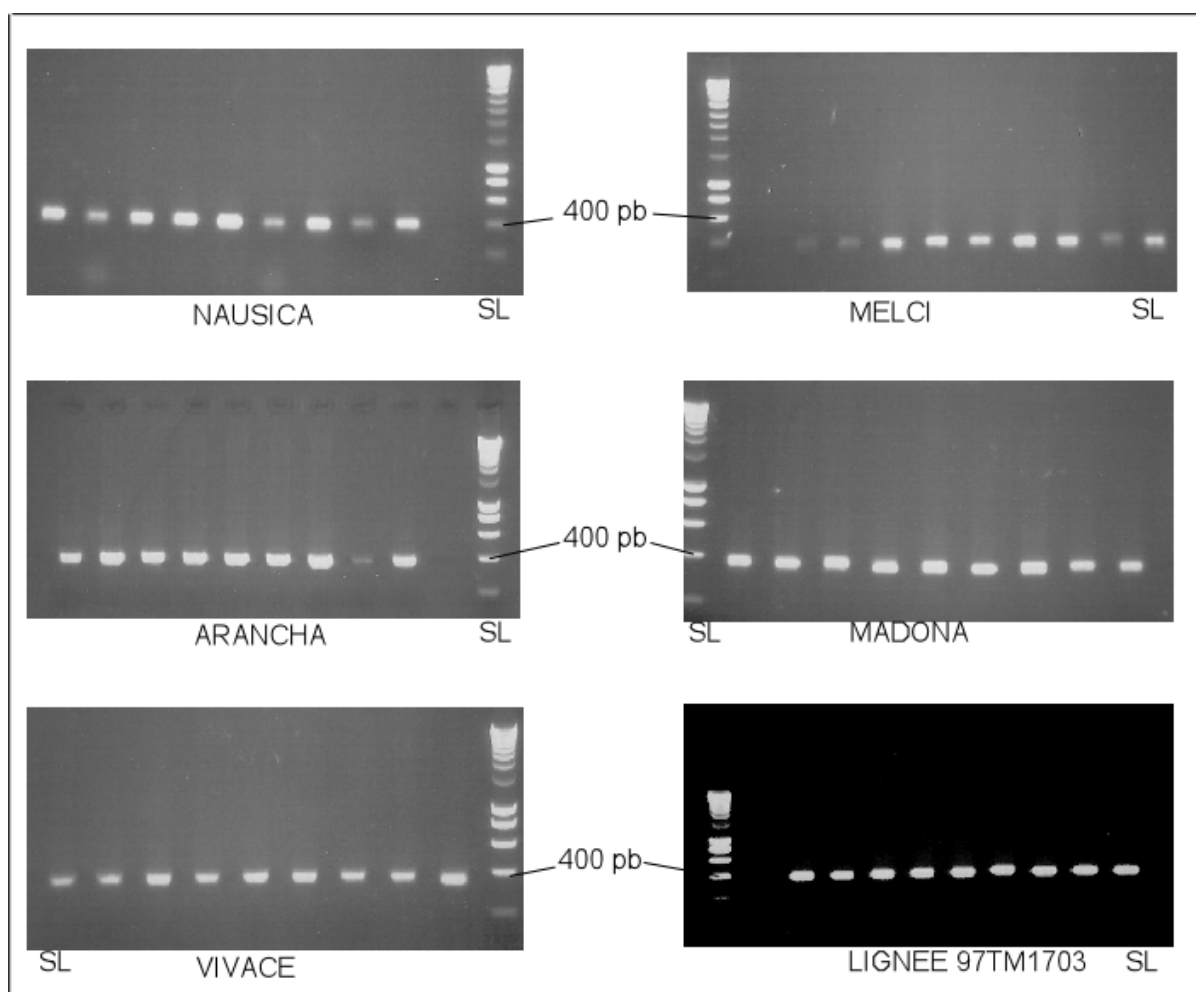


Fig. 12. Migration sur gel d'agarose 2% du produit d'amplification de la paire d'amorces 18 sur l'ADN de neuf individus par variété (taille attendue: 342 pb)

En conclusion, les paires d'amorces 6, 23, 24, 32, 64 et 72 ont été sélectionnées et couplées à un fluorochrome Wellred permettant la détection sur le système d'analyse capillaire (Beckman).

3. Analyse du polymorphisme intravariétal et intervariétal

Afin d'obtenir des informations au sujet de l'organisation génotypique à l'intérieur et entre les 6 variétés de chicorée industrielle, l'ADN d'individus appartenant à ces variétés a été extrait. Ensuite, à l'aide des amorces mises au point, la taille des allèles a été déterminée pour chaque individu. Enfin, les résultats obtenus ont été analysés à l'aide d'un programme statistique.

3.1. Extraction d'ADN

Neuf individus ont été sélectionnés dans chacune des 6 variétés suivantes :

- Nausica
- Melci
- Arancha
- Madona
- Vivace
- 97TM17-03 (lignée consanguine)

Pour chacun des 54 individus, 100 mg de racine broyée ont été traités afin d'en extraire l'ADN (*voir protocole 1.*).

Afin de vérifier la qualité de l'ADN purifié, chaque échantillon a été amplifié à l'aide de la paire d'amorces 18. Bien qu'elle n'ait pas été sélectionnée (en effet, la taille des produits d'amplification ne semblait pas varier d'une variété à l'autre (Figure 11.)), ces amorces ont été choisies pour ce test car elles donnent de très bonnes amplifications.

Pour chaque variété, on observe un produit d'amplification pour la taille attendue (342 pb) pour les neuf échantillons d'ADN (Fig.12). L'ADN est donc de qualité et en quantité suffisante pour pouvoir être amplifié dans les étapes suivantes.

Sur gel d'agarose, les bandes parasites observées sur gel d'acrylamide (Fig.11) ne sont plus visibles. Pourtant, les conditions d'amplification sont identiques et la coloration se fait, dans les deux cas, au bromure d'éthidium. Aucune explication n'est proposée pour cette observation.

3.2. Détermination de la taille des allèles

Afin de déterminer avec précision la taille des deux allèles de chaque individu, quatre étapes sont nécessaires (*matériel et méthode 8.*):

- Test préliminaire des 6 paires d'amorces marquées

Deux échantillons d'ADN ont été sélectionnés et amplifiés à l'aide des 6 paires d'amorces (paires d'amorces 6, 23, 24, 32, 64, 72) couplées à un fluorochrome. Seules

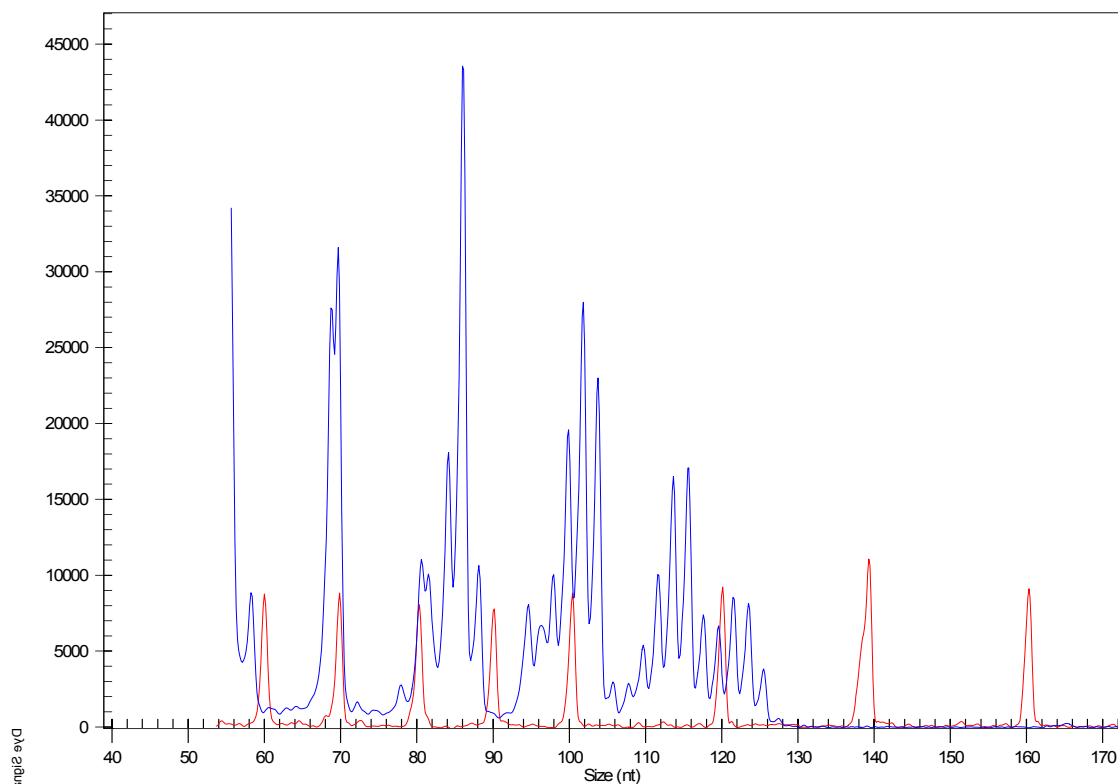


Fig. 13. Graphique représentant la fluorescence émise par un fluorochrome couplé à un fragment microsatellite. Axe des abscisses : taille des fragments. Axe des ordonnées : intensité du signal fluorescent. Ce graphique a été obtenu à l'aide des amorces 24. L'amplification n'est pas spécifique.

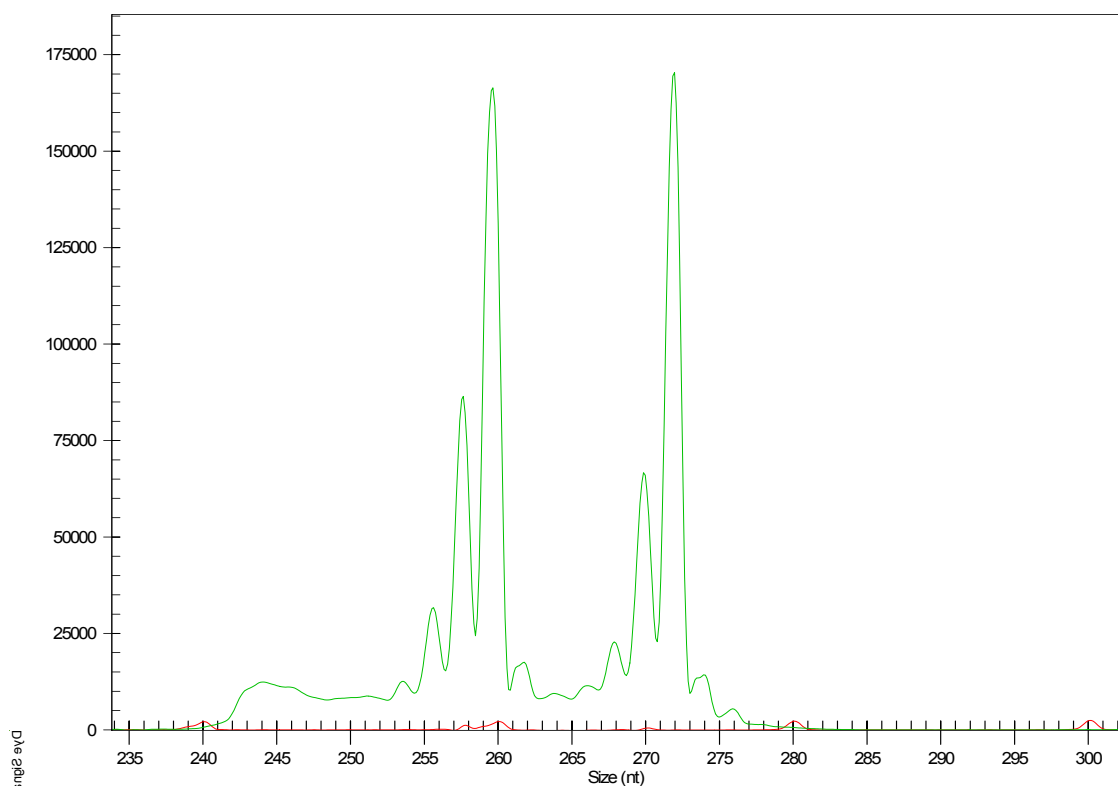


Fig. 14. Graphique représentant la fluorescence émise par un fluorochrome couplé à un fragment microsatellite. Axe des abscisses : taille des fragments. Axe des ordonnées : intensité du signal fluorescent.

les amorces donnant un produit d'amplification pouvant être détecté, avec un profil typique de microsatellite et sans bandes parasites ont été utilisées pour amplifier la totalité des échantillons.

Seules les paires d'amorces 23, 32 et 64 ont été conservées. En effet, aucune amplification n'était détectable pour la paire d'amorces 72 et des amplifications non spécifiques furent observées pour les amorces 6 et 24 (Figure 13).

- Amplification de la totalité des échantillons avec les 3 paires d'amorces microsatellites

Tous les échantillons (soit 54) ont été amplifiés avec les trois paires d'amorces. Au total, 162 PCR ont été effectuées. Pour certains échantillons d'ADN, l'amplification n'a pas eu lieu. Pour ceux-là, l'amplification a été refaite en augmentant la quantité d'ADN. Dans ces conditions, la totalité des échantillons ont fourni un résultat (sauf pour le locus 32 de l'individu 2 de la variété Nausica).

- Migration du produit d'amplification sur gel de séquençage

Le produit d'amplification migre dans un capillaire contenant un gel d'acrylamide en même temps qu'un standard de taille. A l'extrémité du capillaire, la fluorescence des fragments, émise après excitation de leur fluorochrome, est mesurée par un détecteur. Les fragments sont détectés par ordre croissant de taille. Le standard de taille permet de déterminer la taille exacte des deux allèles du locus microsatellite pour chaque individu.

- Détermination de la taille des allèles

Les figures 14 et 15 montrent deux résultats types d'allèles microsatellites.

La figure 14 présente un individu hétérozygote pour le locus microsatellite. En effet, les deux pics majoritaires sont clairement visibles.

On observe que plusieurs pics d'intensité croissante sont présents pour un seul allèle. Le profil ressemble souvent à ceci : l'intensité des pics augmente jusqu'à un maximum puis chute, ne laissant qu'un pic très bas. Le pic précédant celui de très faible intensité est celui pris en compte pour déterminer la taille de l'allèle. Les différents pics correspondent à des amplifications incomplètes du motif microsatellite. En effet, lors de l'amplification, la Taq polymérase « saute » une ou plusieurs répétitions dans le microsatellite donnant ainsi des fragments de taille différente.

La figure 15 représente un individu homozygote pour le locus microsatellite amplifié par l'amorce 64. C'est donc le pic précédant celui de très faible intensité qui est pris en compte pour déterminer la taille des allèles de cet individu homozygote.

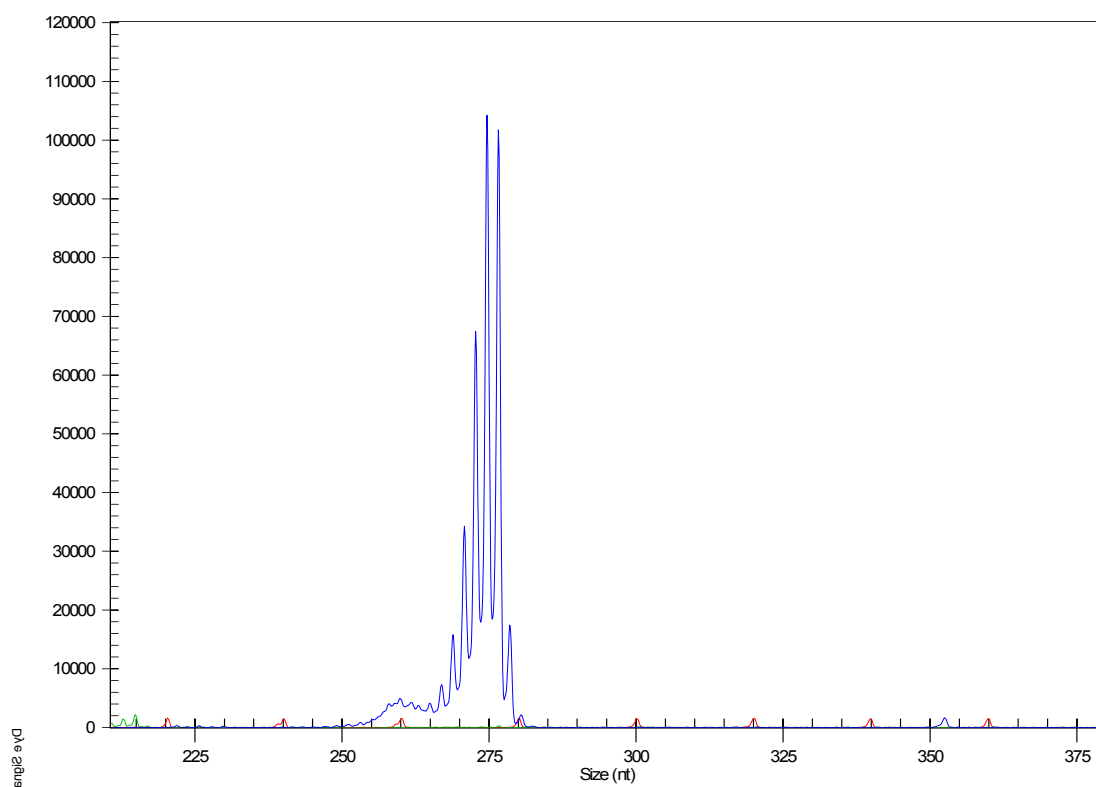


Fig.15 Graphique représentant la fluorescence émise par un fluorochrome couplé à un fragment microsatellite.
Axe des abscisses : taille des fragments. Axe des ordonnées : intensité du signal fluorescent.

NAUSICA						
Amorces:	23	23	64	64	32	32
Ind.1	235	247	278	278	322	322
Ind.2	227	235	276	272	0	0
Ind.3	235	247	274	274	310	322
Ind.4	219	239	272	276	310	310
Ind.5	219	219	276	272	310	310
Ind.6	235	247	266	266	272	272
Ind.7	223	235	276	276	310	310
Ind.8	231	231	266	274	310	322
Ind.9	221	231	278	278	292	310

MADONA						
Amorces:	23	23	64	64	32	32
Ind.1	231	241	272	278	310	310
Ind.2	247	253	266	266	266	322
Ind.3	223	239	276	276	300	300
Ind.4	231	241	272	278	266	304
Ind.5	231	241	266	272	266	304
Ind.6	223	239	276	276	266	304
Ind.7	221	225	276	278	266	266
Ind.8	219	237	272	278	266	310
Ind.9	231	239	272	278	276	310

MELCI						
Amorces:	23	23	64	64	32	32
Ind.1	221	231	270	278	272	310
Ind.2	219	219	270	278	260	310
Ind.3	227	241	278	278	272	322
Ind.4	239	219	270	272	308	323
Ind.5	239	219	270	272	266	266
Ind.6	219	249	270	262	310	322
Ind.7	219	219	270	278	260	260
Ind.8	241	233	270	278	272	272
Ind.9	231	241	262	272	262	310

VIVACE						
Amorces:	23	23	64	64	32	32
Ind.1	219	219	274	278	276	304
Ind.2	219	219	262	272	266	310
Ind.3	247	247	262	262	266	310
Ind.4	219	247	262	272	266	266
Ind.5	219	247	262	272	266	304
Ind.6	247	247	262	262	266	304
Ind.7	219	247	274	278	322	266
Ind.8	219	223	274	278	310	310
Ind.9	219	223	274	278	266	304

ARANCHIA						
Amorces:	23	23	64	64	32	32
Ind.1	259	259	266	266	266	310
Ind.2	219	218	266	272	266	322
Ind.3	219	219	266	272	260	272
Ind.4	219	235	276	272	272	272
Ind.5	245	259	266	266	260	272
Ind.6	219	259	266	272	266	322
Ind.7	219	259	266	272	266	320
Ind.8	215	215	276	276	274	322
Ind.9	219	219	266	278	266	322

LIGNEE 97TM1703						
Amorces:	23	23	64	64	32	32
Ind.1	241	241	268	268	310	310
Ind.2	241	249	270	270	310	310
Ind.3	241	241	268	268	310	310
Ind.4	241	241	268	268	310	310
Ind.5	241	241	268	268	310	310
Ind.6	241	241	268	268	310	310
Ind.7	227	241	268	272	310	322
Ind.8	241	241	268	268	310	310
Ind.9	221	231	282	286	310	310

Tableau 7. Taille des allèles de trois loci microsatellites amplifiés par les amorces 23, 32 et 64 pour 9 individus (Ind.) issus de 6 variétés de chicorée industrielle.

3.3. Interprétation des résultats

Le tableau 7 reprend la taille des deux allèles de chaque locus (23, 64, 32) pour les neuf individus issus des 6 variétés analysées.

Parmi ces 6 variétés, une lignée est présente. Par définition, une lignée est un ensemble d'individus possédant le même patrimoine génétique et étant homozygote pour chacun de ses gènes.

La lignée 97TM17-03 est composée d'individus issus de deux autofécondations successives. La première autofécondation a été réalisée sur la plante mère 97-03 (variété Polanowika). La deuxième autofécondation s'est faite sur un individu de la F1 : le plant 97TM17-07. Ce sont les akènes provenant de l'autofécondation de ce plant qui ont fourni les racines analysées. Le taux d'homozygotie attendu dans la génération F2 est de 75%. Les taux observés (tableau 7) pour les loci 23, 64, 32 sont respectivement de 66,6%, 77,7% et 88,8% soit une moyenne de 77,7%. Les valeurs attendues et observées sont très proches, ce qui confirme que la lignée 97TM17-03 a bien été obtenue suite à deux autofécondations successives.

Afin de retirer un maximum d'informations de ces données, un programme d'analyse de la diversité génétique appelé « SPAGeDi » a été utilisé.

Les paramètres suivants ont été sélectionnés :

- Niveau d'analyse : (2) population
- Statistiques :
 - basées sur l'identité des allèles : (1) F- statistiques globales et pairées
 - basées sur la taille des allèles : (6) R-statistiques globales et pairées
- Options « computational » : (3) faire des tests de permutation
- Options de permutation : (1) tester la structure génétique
(2) tester l'information de la taille des allèles
- Options supplémentaires : (1) détail des tests de permutation
- Nombre de permutations : 5000
- Options de sortie : (3) présenter les matrices avec les coefficients génétiques pairés
- Options sur le format des coefficients génétiques pairés : (-1) écrire en matrice

Les options ainsi déterminées permettent d'obtenir différents résultats :

1. Les fréquences alléliques

Les fréquences alléliques donnent une idée du polymorphisme des différents loci et de la diversité génétique au sein des variétés.

Le tableau 8 montre la fréquence allélique pour chaque locus. Cette donnée se mesure de la façon suivante :

amorce 23 :17 alleles:	215	219	221	223	225	227	231	233	235	237	239	241	245	247	249	253	259
Nausica		3	1	1		1	3		5		1			3			
melci		7	1			1	2	1			2	3			1		
Arancha	2	9							1				1				5
Madona		1	1	2	1		4			1	3	3		1		1	
Vivace		9		2										7			
Lignée97TM1703			1			1	14								1		
TOTAL	2	29	4	5	1	3	23	1	6	1	6	6	1	11	2	1	5
Fréquence allélique	0.019	0.259	0.037	0.046	0.009	0.028	0.212	0.009	0.056	0.009	0.056	0.056	0.009	0.102	0.019	0.009	0.046

amorce 64 :10 alleles	262	266	268	270	272	274	276	278	282	286
Nausica		3			3	3	5	4		
Melci	2			7	3		6			
Arancha		9			5		3	1		
Madona		3			5		5	5		
Vivace	7				3	4		4		
Lignée 97TM1703			13	2	1				1	1
TOTAL	9	15	13	9	20	7	19	14	1	1
Fréquence allélique	0.083	0.139	0.12	0.083	0.185	0.065	0.175	0.129	0.009	0.009

amorce 32:14 alleles	260	262	266	272	274	276	292	300	304	308	310	320	322
Nausica				2			1				9		4
Melci	3	1	2	4						1	4		3
Arancha	2		5	4	1						1	1	4
Madona			7			1		1	3		4		1
Vivace			9			1			4		4		1
Lignée 97TM1703											17		1
TOTAL	5	1	23	10	1	2	1	1	7	1	39	1	14
Fréquence allélique	0.047	0.009	0.212	0.092	0.009	0.019	0.009	0.009	0.066	0.009	0.368	0.009	0.123

Tableau 8. Fréquences de chaque allèle obtenu par les paires d'amorces 23, 32 et 64 dans les six variétés

Soit un locus biallélique Aa

Génotype de l'échantillon : AA : X individus
 Aa : Y individus
 aa : Z individus

Fréquence allélique de A : $2X+Y$ (nombre d'allèles A) / $2(X+Y+Z)$ (nombre total d'allèles)

Le concept de fréquence allélique sous-entend l'idée du polymorphisme. Un gène polymorphe est un gène pour lequel l'allèle le plus commun a une fréquence inférieure à 0,95 (D.L. Hart, 1994). Dans le cas de nos trois loci, aucune fréquence allélique n'est supérieure à 0,95. On peut donc dire que nos trois loci sont polymorphes.

En plus d'être polymorphes, nos trois loci présentent un grand nombre d'allèles (respectivement 17, 14 et 10). Ce qui laisse supposer que la diversité génétique chez les variétés de chicorée industrielle est importante.

2. Les F-Statistiques

Les F-statistiques mesurent les relations génétiques intra- et inter-variétales. Elles permettent de donner une approximation de la consanguinité au sein des variétés ainsi qu'une estimation de la divergence intervariétale.

Trois coefficients sont repris dans les F-statistiques (D.L. Hart, 1994) :

- Le **Fis** rend compte de la consanguinité au sein des variétés (variabilité intravariétale). Il représente donc la diminution d'hétérozygotes due à ce facteur dans chaque variété.
- Le **Fst** correspond à la séparation génétique entre les variétés (variabilité intervariétale). Il représente donc la diminution d'hétérozygotes due à ce facteur.
- Le **Fit** est le coefficient global de consanguinité. Il reprend les effets de la consanguinité intravariétale (Fis) et de la subdivision intervariétale (Fst).

Voici les F-statistiques globales calculées **pour les 6 variétés** de manière générale et pour chaque locus séparément :

F-statistiques globales			
	Fit	Fis	Fst
Tous les loci	0.2903	0.1274	0.1867
CiGa23	0.2598	0.0943	0.1827
CiGa32	0.3232	0.1525	0.2014
CiGa64	0.2876	0.1365	0.175

Fst :

Une valeur de 0 signifie qu'il n'y a aucune séparation génétique entre les variétés. La valeur moyenne obtenue du Fst (0,19) signifie qu'il y a une grande différence génétique entre les variétés. Cela indique que les variétés sont génétiquement bien séparées les unes des autres. Si l'on regarde chaque locus séparément, les conclusions sont identiques car les valeurs sont comprises entre 0,175 et 0,201. Ce qui correspond toujours à une grande différenciation génétique. Les similarités des valeurs observées démontrent la cohérence des informations fournies par chaque locus séparément.

Fis :

Une valeur nulle signifie qu'il n'y a aucune consanguinité entre les individus et donc que la reproduction se fait de manière totalement aléatoire. La valeur moyenne obtenue (0,12) signifie que la consanguinité au sein des variétés n'est pas nulle et donc que la reproduction entre les individus ne se fait pas tout à fait de manière aléatoire. L'effet de la consanguinité est une augmentation du taux d'homozygotie. Les conclusions sont identiques si l'on analyse les Fis pour chaque locus et comme le Fst, les similarités des valeurs observées démontrent la cohérence des informations fournies par chaque locus séparément.

La présence de consanguinité pourrait être expliquée par la sélection artificielle. Supposons que cette variété soit issue du croisement de deux individus, le pool d'allèles de départ va donc être restreint. Les allèles identiques au sein de la variété seront donc plus nombreux que s'il n'y avait pas eu sélection. Ce phénomène implique la présence de consanguinité. Cette hypothèse ne peut être vérifiée : l'origine des variétés étudiées n'est pas connue.

Fit :

Cette valeur est proche de la somme des effets de la consanguinité et de la séparation entre les variétés : $0.1274 \text{ (Fis)} + 0.1867 \text{ (Fst)} \approx 0.2903 \text{ (Fit)}$. Ce coefficient permet d'avoir une vue générale sur l'augmentation du taux d'homozygotie (ou la diminution du taux l'hétérozygotie).

La lignée 97TM17-03 est issue de deux autofécondations successives. Cela implique que le taux d'homozygotie dû à la consanguinité est élevé (75%). On suppose que la valeur du Fis globale (0,12) est principalement due à cette variété.

Ainsi, afin de savoir si la lignée a un effet sur le calcul du coefficient de consanguinité (Fis) et sur la séparation génétique (Fst), la même analyse a été effectuée mais **sans les données de la lignée 97TM17-03** :

F-statistiques globales			
	Fit	Fis	Fst
Tous les loci	0.191	0.1085	0.0925
CiGa23	0.1681	0.0849	0.0909
CiGa32	0.2031	0.0998	0.1148
CiGa64	0.2023	0.1406	0.0718

Fis :

Cette valeur diminue globalement de 15% lorsque les données de la lignée 97TM17-03 sont supprimées. Le logiciel utilisé (SPAGeDI) ne donne pas le détail par variété. Il serait utile de disposer d'un logiciel fournissant ces informations afin de pouvoir interpréter cette diminution.

Fst

Cette valeur, par contre, diminue de moitié. Une valeur globale de 0,09 correspond à une séparation génétique modérée entre les variétés. Nous pouvons donc faire l'hypothèse que la grande séparation génétique observée pour les 6 variétés est due principalement à la présence de la lignée autofécondée.

En ce qui concerne l'aspect général des données, on peut observer une certaine concordance des Fis, Fst et Fit d'un locus à l'autre. Cette observation permet d'appuyer l'hypothèse que ces valeurs sont cohérentes.

En conclusion, les 5 variétés sont séparées génétiquement l'une de l'autre de manière modérée et la consanguinité est présente au sein de chacune d'elles.

3. Les R-statistiques

Les R-statistiques sont l'équivalent des F-statistiques à la différence qu'elles prennent en compte la taille des allèles. En effet, ce modèle se base sur l'hypothèse que deux allèles de tailles voisines sont plus proches que deux allèles de tailles éloignées. Cela signifie que chaque variété posséderait sa propre gamme de tailles pour un locus donné. Pour que ce phénomène soit observable, il faut que les variétés soient séparées depuis un temps suffisamment important. Dans le cas où les variétés ne sont pas suffisamment éloignées pour posséder une gamme de tailles alléliques spécifique, les R-statistiques ne donneront pas de résultats significatifs (O. Hardy & X. Vekemans, 2002).

R-statistiques Globales			
	Rit	Ris	Rst
Tous les loci	0.1614	0.0093	0.1535
CiGa23	0.3151	0.3088	0.0092
CiGa32	0.5226	0.5004	0.0444
CiGa64	0.0981	-0.1283	0.2007

Le tableau ci-dessus montre des valeurs extrêmement variables. Comparées aux valeurs observées pour les F-statistiques (page précédente), aucune homogénéité n'est observable entre les valeurs des différents loci. Il semblerait que le bruit de fond soit très important. Avant d'analyser ces résultats, il est donc nécessaire de vérifier que ces valeurs peuvent être prises en compte.

4. Tests de permutations

Deux types de tests de permutations ont été réalisés dans des buts différents :

- Test de permutations des individus et des allèles : ce test permet de vérifier que les F-statistiques et R-statistiques sont significativement différentes d'une distribution statistique normale.

- Test de permutations de la taille des allèles : ce test permet de déterminer quels paramètres (mutation et/ou dérive) influencent la divergence génétique.

a) Test de permutations des individus et des allèles

Pour rappel (O. Hardy & X. Vekemans, 2002) :

- $F-R_{it}$ représente la corrélation de deux allèles d'un **individu** (i) par rapport à deux allèles pris au hasard dans la **population totale** (t).
- $F-R_{is}$ représente la corrélation de deux allèles d'un **individu** (i) par rapport à deux allèles pris au hasard dans une **sous-population** (s).
- $F-R_{st}$ représente la corrélation entre deux allèles pris dans une **sous-population** (s) par rapport à deux allèles pris au hasard dans la **population totale** (t).

Le test de permutation donne 5000 valeurs des F- ou R- statistiques obtenues lors de 5000 permutations aléatoires :
 - des allèles dans la population totale pour le $F-R_{it}$,
 - des allèles dans la variété pour le $F-R_{is}$,
 --des individus dans la population totale pour le $F-R_{st}$.

Pour chaque coefficient, les 5000 valeurs se répartissent suivant une distribution normale dont la moyenne des valeurs obtenues est positionnée au point 0.

L'étape suivante est de voir si la valeur observée (celle obtenue sans permutation) est significativement différente ou non de la valeur attendue (celle obtenue avec permutations).

Deux hypothèses sont alors émises :

- **Hypothèse nulle (H_0)** : la valeur attendue est égale à la valeur observée soit $obs = exp$
- **Hypothèse alternative (H_1)** : la valeur attendue est différente de la valeur observée soit $obs \neq exp$.

Si la probabilité (P) que la valeur observée se trouve dans la distribution normale est inférieure à 0,05, alors H_0 rejetée et H_1 est acceptée. Dans ce cas, cela signifie que notre valeur observée est significativement différente d'une distribution statistique aléatoire et donc que nous pouvons la prendre en compte.

Par contre si la probabilité (P) que la valeur observée se trouve dans la distribution normale est supérieure à 0,05, alors H_0 est acceptée et H_1 rejetée. Cela signifie que la valeur observée n'est pas significativement différente d'une distribution statistique aléatoire et donc que nous ne pouvons la prendre en compte.

Les probabilités P pour les **F-statistiques** sont les suivantes :

F-statistiques globales			
	P pour F_{it}	P pour F_{is}	P pour F_{st}
Tous les loci	0	0.0044	0
CiGa23	0	0.1542	0
CiGa32	0	0.07	0
CiGa64	0	0.076	0

Certaines probabilités pour les Fis sont supérieures à 0,05 mais en moyenne (« tous les loci »), H_0 peut être rejetée. Fit, Fis et Fst sont donc considérés comme significativement différents d'une distribution statistique aléatoire. Ces valeurs peuvent donc être prises en compte.

Les probabilités P pour les **R-statistiques** sont les suivantes :

	R-statistiques globales		
	P pour Rit	P pour Ris	P pour Rst
Tous les loci	0.1342	0.9376	0
CiGa23	0.0194	0.0648	0.7251
CiGa32	0	0.001	0.3889
CiGa64	0.4769	0.4361	0

Seul le Rst moyen peut être pris en compte. Pour ce qui concerne les autres valeurs, elles ne sont pas significativement différentes d'une distribution statistique aléatoire puisque l'hypothèse nulle n'est pas rejetée. De manière générale, il semblerait donc que le bruit de fond dû au paramètre « taille des allèles » soit très important lors du calcul des R-Statistiques. Cela signifie que les variétés ont été définies depuis trop peu de temps pour posséder leur propre gamme de tailles d'allèles.

b) Test de permutation de la taille des allèles

Ce test a pour but de déterminer quels facteurs influencent la divergence génétique présente entre les variétés. Le principe est similaire au test de permutation présenté ci-dessus à la différence que ce ne sont plus, ni les allèles, ni les individus qui sont permutés mais la taille des allèles (O. Hardy & X. Vekemans, 2002).

Afin de comprendre le principe de ce test, certains concepts doivent être expliqués :
Le Fst et le Rst peuvent être expliqués par deux formules :

Fst	Rst
$\frac{1}{1+4N_e(m+\mu)}$	$\frac{1}{1+4 N_e m}$

En effet, la variation intervariétale est influencée par les paramètres suivants (D.L. Hart, 1994):

- La dérive génétique (N_e) : elle représente la dispersion progressive des fréquences alléliques parmi les variétés et l'augmentation simultanée du nombre de variété fixées (D.L. Hart, 1994). Cela signifie plus simplement que les allèles deviennent petit à petit spécifiques d'une variété car il n'y a plus ou peu de contact avec les autres variétés. Cela implique donc que la différence génétique entre les variétés est de plus en plus importante.

La dérive génétique est exprimé par « N_e ». « N_e » est l'effectif efficace de la population c'est-à-dire sa taille. Plus celle-ci est faible et plus la divergence entre les variétés sera grande.

- Le flux de gènes (m) : le flux de gènes est aussi appelé migration. La migration retarde la divergence entre les variétés puisqu'il y a échange d'allèles d'une variété à l'autre. « m » représente le nombre de migrants.
- Le taux de mutations (μ) : les effets cumulés des mutations sur une longue période peuvent influencer les fréquences alléliques d'une variété. Par contre, sur une période de quelques générations, les modifications des fréquences alléliques dues aux mutations seront très faibles.

Dans le Rst, le taux de mutation n'est plus pris en compte. Rst est donc toujours plus grande ($\mu > 0$) ou égale ($\mu = 0$) au Fst.

Lors du test de permutation de la taille allélique, la valeur obtenue est plus ou moins l'équivalent du Fst. En effet, le fait de permuter la taille des allèles de manière aléatoire a pour conséquence de donner une valeur ne tenant plus compte de ce paramètre et donc correspondant au Fst (le Fst n'est pas influencé par la taille des allèles).

Deux cas sont possibles :

- La valeur attendue (\approx Fst) est $<$ à la valeur observée (Rst). Dans ce cas, cela signifie que le taux de mutation a une influence sur la divergence génétique.
- La valeur attendue (\approx Fst) est $=$ à la valeur observée (Rst). Cela signifie que le taux de mutation n'a pas d'influence sur la divergence génétique.

Les résultats obtenus pour le test de permutation de la taille des allèles sont les suivants :

Valeur attendue (\approx Fst)	0.170875
Valeur observée ($=$ Rst)	0.153521

Dans notre cas, la valeur observée est légèrement inférieure à la valeur attendue. On suppose dès lors que le Fst et le Rst sont identiques. Cela signifie donc que le taux de mutation (μ) n'a pas d'influence sur la divergence génétique de nos variétés car celles-ci sont séparées depuis un temps trop court. La divergence génétique est donc due principalement à la dérive génétique (exprimé par N_e)

En résumé, les deux test de permutation nous apportent les informations suivantes:

- Les R-statistiques ne sont pas significativement différentes d'une distribution statistique aléatoire et donc, ces valeurs ne peuvent être prises en compte. On en conclut que la taille des allèles ne donne pas d'information supplémentaire. Cela implique que la prise en compte de la taille de l'allèle lors du calcul du Rst introduit un important bruit de fond, ce qui fausse les résultats.

- Le taux de mutation n'influence pas la divergence génétique entre les variétés. C'est la dérive génétique qui est le principal responsable de cette divergence.

En conclusion, le fait que les variétés ne possèdent pas leur propre gamme de tailles alléliques et que le taux de mutation n'influence pas la divergence génétique montre que les variétés sont séparées l'une de l'autre depuis un temps relativement court.

5. Distances génétiques pairées

Le calcul de la distance génétique pairée permet de déterminer la différenciation génétique entre deux variétés.

Les F_{st} sont donc calculés pour les variétés deux à deux :

	Nausica	Melci	Arancha	Madona	Vivace	97TM1703
Nausica		0.071	0.101	0.046	0.127	0.298
Melci	0.071		0.096	0.068	0.102	0.326
Arancha	0.101	0.096		0.086	0.142	0.434
Madona	0.046	0.068	0.086		0.097	0.328
Vivace	0.127	0.102	0.142	0.097		0.434
97TM1703	0.299	0.326	0.434	0.328	0.434	

Pour rappel (D.L. Hart, 1994) : F_{st} :

- entre 0 et 0,05 : petite différenciation génétique
- entre 0,05 et 0,15 : différenciation génétique modérée
- entre 0,15 et 0,25 : grande différenciation génétique
- $> 0,25$: très grande différenciation génétique

Nous pouvons tirer les informations suivantes de ce tableau :

- Madona semble avoir le moins de différenciation génétique avec les autres variétés.
- La lignée 97TM17-03 est la variété la plus divergente.
- Nausica et Melci sont relativement proches génétiquement l'une de l'autre.
- Vivace et Arancha semblent plus éloignées génétiquement des autres variétés.

Même si certaines variétés semblent plus ou moins différenciées l'une de l'autre, la différenciation globale entre les 5 variétés (sans la lignée) reste tout de même modérée (F_{st} globale : 0,0925). Par contre, la lignée 97TM17-03 présente une très grande différenciation génétique par rapport aux cinq autres variétés (F_{st} pairée : de 0,2985 à 0,4339).

4. Discussion générale

L'objectif de ce mémoire était de réaliser une étude de la diversité génétique intra- et inter-variétale au sein de six cultivars de chicorées industrielles (*Cichorium intybus*). Dans ce but, il a fallu procéder au choix d'un type de marqueur moléculaire.

Bien qu'historiquement les techniques RFLP et RAPD aient été largement utilisées, elles ont depuis été supplantées par l'émergence de techniques telles que l'AFLP, les microsatellites et plus récemment, les SNP.

Bien que nécessitant une mise au point plus importante que pour l'AFLP, les marqueurs microsatellites se sont vite imposés comme étant des outils de choix pour notre étude. A l'opposé de l'AFLP, les microsatellites sont des marqueurs co-dominants, bénéficiant d'une reproductibilité parfaite et pouvant être utilisés, une fois les mises au point effectuées, sur un grand nombre d'individus en un temps restreint. L'absence d'étape de restriction et/ou ligation en fait également une technique robuste, beaucoup moins sensible à d'éventuels contaminants présents au niveau de l'ADN analysé. De leur côté, les SNP ont, dès le début, été rejetés puisque nécessitant l'accès à des banques de données issues d'un séquençage systématique ou du séquençage d'EST. A l'heure actuelle, de telles banques de données sont seulement en cours d'élaboration pour la chicorée industrielle.

Les principales limitations des marqueurs microsatellites proviennent des nombreuses étapes nécessaires à leur sélection. Au cours de ce mémoire, nous avons pu réaliser l'ensemble de ces étapes, aboutissant, au final, au développement de trois marqueurs utilisables.

Les résultats des différentes étapes mises en œuvres ont été présentés dans le chapitre précédent. Il nous paraît cependant opportun de revenir un peu plus en détail sur l'étape clé du développement de marqueurs microsatellites, à savoir l'étape d'enrichissement en séquences microsatellites de la collection de fragments d'ADNg par hybridation soustractive (sonde GA₁₅).

Cette étape dérive de la méthode d'isolement de séquences microsatellites proposée par Kandpal *et al.*, en 1994.

Cette étape vise à isoler des fragments d'ADN simple brin contenant un motif microsatellite complémentaire à celui d'une sonde nucléotidique biotinylée de séquence connue. La séquence de cette sonde est, ici, GA₁₅. Le choix d'un dinucléotide permet d'obtenir des marqueurs neutres. En effet, les répétitions dinucléotidiques sont beaucoup moins fréquentes dans les régions codantes que les microsatellites trinucléotidiques. De plus, les répétitions GA semblent beaucoup plus nombreuses dans le génome que les autres types de microsatellites (K.J. Edwards *et al.*, 1995).

Le criblage des clones s'est fait à l'aide d'une sonde biotinylée. Cette technique permet d'employer les mêmes sondes que celles utilisées lors de l'hybridation soustractive. De plus, l'utilisation de la révélation à l'aide de la phosphatase alcaline et d'un substrat chimioluminescent permet d'éviter l'utilisation d'un marquage radioactif avec tous les inconvénients que cela comporte (mesures de sécurité importantes, infrastructures spéciales, réactifs instables,...)

Suite au premier essai d'enrichissement, réalisé sur base d'un protocole "standard", il est vite apparu après clonage, repiquage et criblage froid par Southern, que les conditions définies

nécessitaient une sérieuse optimisation puisque seuls quatre clones sur 100 se sont avérés être positifs lors du criblage.

Après avoir consulté nombre de protocoles relatifs à cette étape d'enrichissement, il est apparu qu'il n'existait pas de consensus évident au niveau des paramètres principaux, à savoir : température d'hybridation, température, conditions ioniques et nombre des lavages. Cependant, il semblerait que certains protocoles utilisent une température d'hybridation inférieure de 17,5°C au T_m théorique. Dans le second protocole, la température d'hybridation a donc été augmentée à 65°C. La concentration en sel a également été modifiée afin d'augmenter la stringence des lavages.

Après analyse des clones obtenus, il apparaît qu'aucun clone isolé ne contient de fragment possédant des motifs microsatellites (aucun clone ne répond en Southern, mis à part le contrôle positif). Nous avons donc supposé que les conditions étaient trop stringentes. Le troisième protocole a donc été basé sur le premier, tout en modifiant quelques paramètres tels que les conditions, le nombre et la température des lavages, et cela de manière modérée. Dans ces conditions, le rendement a pu être augmenté jusqu'à 4x (56 clones sur 350 étaient positifs lors du criblage sur membrane).

Après séquençage et analyse d'une cinquantaine de clones, les amorces microsatellites ont donc pu être définies. Douze paires d'amorces ont été sélectionnées sur base de deux critères : minimiser les interactions au sein (structures secondaires) et entre les amorces et sélectionner les séquences comprenant le plus grand nombre de répétitions afin d'augmenter les chances de polymorphisme.

Les conditions d'amplification de ces amorces ont été optimisées. L'utilisation d'un appareil PCR possédant un gradient de température a permis de trouver rapidement la température à laquelle un maximum de paires d'amorces amplifiait.

Ensuite, afin de voir si les amplicons présentaient du polymorphisme au sein des différentes variétés, ceux-ci ont été analysés sur gel d'acrylamide 12%. Cette étape permet de voir la variation de la taille des bandes. Cependant, un nombre important de bandes parasites dont la présence n'est pas expliquée a été visualisé. Afin de remédier à ce problème, les produits d'amplification ont été dénaturés avant la migration, mais sans résultat. Toutefois, lors de la migration de ces amplicons sur gel d'agarose 2%, les bandes parasites ne sont plus visibles. Il serait donc intéressant d'optimiser les conditions de migration sur gel d'agarose afin de visualiser les variations de taille des séquences amplifiées.

Six paires d'amorces dont les amplicons semblaient présenter du polymorphisme entre les variétés ont été sélectionnées afin d'être couplées à un fluorochrome. Dans le but de réaliser des multiplexages, deux fluorochromes de longueurs d'onde d'émission différentes ont été choisis. Ainsi, les amorces dont la taille des amplicons était proche ont été couplées avec des fluorochromes différents. Ces deux produits d'amplification pouvaient donc migrer dans le même capillaire tout en étant facilement distinguables.

Au final, trois amorces ont été retenues. En effet, sur les six, une paire d'amorces ne donnait pas d'amplification et deux autres présentaient un nombre important de bandes non spécifiques. Les trois paires d'amorces restantes ont été utilisées sur six variétés différentes de chicorée industrielle : Nausica, Melci, Arancha, Madona, Vivace et la lignée consanguine 97TM17-03. Les résultats obtenus ont permis de tirer différentes conclusions en ce qui concerne l'organisation génétique intra- et intervariétale de la chicorée industrielle.

Les trois marqueurs microsatellites (23, 32 et 64) présentent du polymorphisme.

Sur base des tailles d'allèles déterminées lors de l'analyse sur l'analyseur de fragments Beckman, il est maintenant possible de déterminer le nombre de répétitions correspondant à chacun des allèles.

Pour rappel, l'isolement des séquences microsatellites s'est fait au départ d'ADN provenant d'un individu issu de la lignée 97TM17-03. Ce individu ne fait pas partie des 9 échantillons analysés.

Avec la paire d'amorce 32, 17 des 18 allèles isolés pour les 9 individus issus de la lignée, ont une taille de 310 bp. Or, sur base du séquençage, la taille attendue est de 301 bp. Cette différence de 9 bp s'explique aisément. Pour des raisons pratiques, une séquence de 7 nucléotides (GTTTCTT) a systématiquement été ajoutée à l'extrémité 5' de l'amorce reverse. En effet, d'après Brownstein *et al* (1996), l'adjonction de cette séquence favorise l'ajout d'une adénine à l'extrémité 3' du fragment amplifié, permettant de cette manière de distinguer le pic principal des "stutter bands" (produits d'amplifications incomplètes). De plus, la firme Beckman signale qu'un décalage de 2 à 3 paires de bases peut exister entre la taille réelle et la taille estimée.

Pour la paire d'amorces 32, les tailles amplifiées varient entre 260 et 322 bp (tailles estimées). Ces tailles correspondent à un nombre de répétitions allant de 7 à 38 (en considérant la taille estimée "310" comme étant identique au fragment réel "301" présentant 32 répétitions). On peut donc observer une grande variabilité du nombre de répétitions présentes au sein de ce motif microsatellite.

En se basant sur la démarche adoptée pour déterminer la correspondance fragments amplifié vs fragments détecté, pour la paire d'amorce 23, une taille de 201 bp ($192 + 9$) correspond à un fragments possédant 21 répétitions GA. Dès lors, on observe entre 28 et 50 répétitions GA parmi les 54 individus analysés (215 à 259 bp).

Enfin, l'amorce 64 voit ses allèles distribués entre 262 et 286 bp, ce qui correspond à un nombre de répétitions allant de 39 à 51. Pour rappel, le fragment attendu sur base du séquençage était de 217 bp, ce qui, majoré des 9 bp donne un fragment de 226 bp correspondant à 21 répétitions.

Le grand nombre d'allèles pour chaque marqueur (respectivement 17, 10 et 13) ainsi que le taux élevé d'hétérozygotie (0.88) laissent supposer que la diversité au sein des variétés de chicorée industrielle est relativement importante.

La lignée issue de deux autofécondations successives contient en moyenne 77,7% de loci homozygotes. Cette valeur observée est très proche de la valeur attendue (75%). Cela confirme le stade d'autofécondation de deuxième génération de cette lignée.

Les tests de permutation des individus et des allèles ont montré que les F-statistiques obtenues sont significativement différentes d'une distribution statistique aléatoire. Elles ont donc pu être prises en compte.

Le calcul païré de la distance génétique ainsi que le calcul du Fst global pour les cinq variétés indiquent que les variétés Nausica, Melci, Arancha, Madona et Vivace sont différenciées génétiquement l'une de l'autre de manière **modérée** (Fst compris entre 0,05 et 0,15). Par

contre, la lignée 97TM17-03 montre une **très grande** différenciation génétique par rapport aux cinq autres variétés (F_{st} supérieur à 0,25).

Le F_{is} (coefficient de consanguinité) diminue d'un sixième lorsque celui-ci est calculé sans tenir compte des valeurs de la lignée. L'utilisation d'un logiciel permettant de détailler les contributions de chaque population, ce qui n'est pas réalisable avec SPAGeDI, nous permettra de mieux percevoir les raisons de la modification observée.

Le test de permutation des individus et des allèles n'a pas pu définir les R-statistiques obtenues comme significativement différentes d'une distribution statistique aléatoire. Ces valeurs n'ont donc pas pu être interprétées. On peut expliquer ce phénomène par le fait que les R-Statistiques prennent en compte la taille des allèles. Il semblerait que, dans notre cas, ce paramètre n'apporte pas d'information supplémentaire, voire même qu'il introduit une source de bruit de fond importante dans les résultats. L'hypothèse émise est que les variétés ont été définies depuis trop peu de temps pour pouvoir posséder leur propre gamme de tailles d'allèles. La taille n'apporte donc aucune information puisqu'elle n'est pas spécifique d'une variété.

Le test de permutation de la taille des allèles permet d'évaluer la contribution apportée par le taux de mutation et par la dérive génétique sur la différenciation génétique des variétés. Pour cela, la valeur obtenue lors du test (correspondant plus ou moins au F_{st} global) a été comparée avec le R_{st} global. Dans notre cas, ces deux valeurs ont été définies comme identiques. Cela signifie que le taux de mutation (μ) n'intervient pas dans la divergence génétique des variétés. Le facteur principalement responsable de la différenciation est donc la dérive génétique (N_e).

Le fait que les variétés ne possèdent pas leur propre gamme de tailles alléliques et le fait que le taux de mutation n'influence pas la divergence génétique montrent que les variétés sont séparées l'une de l'autre depuis un temps relativement court.

Suite aux résultats obtenus lors des analyses des F-statistiques, il nous a paru intéressant de procéder à des analyses de distance afin de pouvoir utiliser ces résultats pour la réalisation de dendrogrammes. Les analyses de distance ont été menées à deux niveaux distincts: de manière globale, entre variétés, et de manière détaillée, en comparant les individus deux à deux.

Pour obtenir les matrices de distances, nous avons utilisé le logiciel "microsat analyser". Ce logiciel permet la génération de matrices de distances au départ de données microsatellites.

Face à la diversité des méthodes de calcul de distance proposées, nous avons décidé d'utiliser la méthode dite de "proportion d'allèles partagés" (Jin and Chakraborty, 1993). De plus amples détails concernant l'algorithme utilisé pour ce calcul sont disponibles à l'adresse suivante : <http://www.pge.cnrs-gif.fr/bioinfo/populations/formules/node3.html>.

Le choix de cette méthode a été fait sur conseil du Pr. Bertin (UCL). Lors d'une étude réalisée au sein de son laboratoire sur des variétés d'orge dont le pedigree était bien établi, il est apparu empiriquement (essai de l'ensemble des méthodes proposées par le logiciel microsat analyser) que c'était cette méthode qui fournissait les résultats les plus proches de la réalité. Nous avons dès lors également opté pour cette méthode.

Les matrices obtenues sont reprises en annexe 1 et 2. Le logiciel en ligne "webPHYLIP" a été utilisé pour analyser ces résultats (<http://biocore.unl.edu/WEBPHYLIP/>).

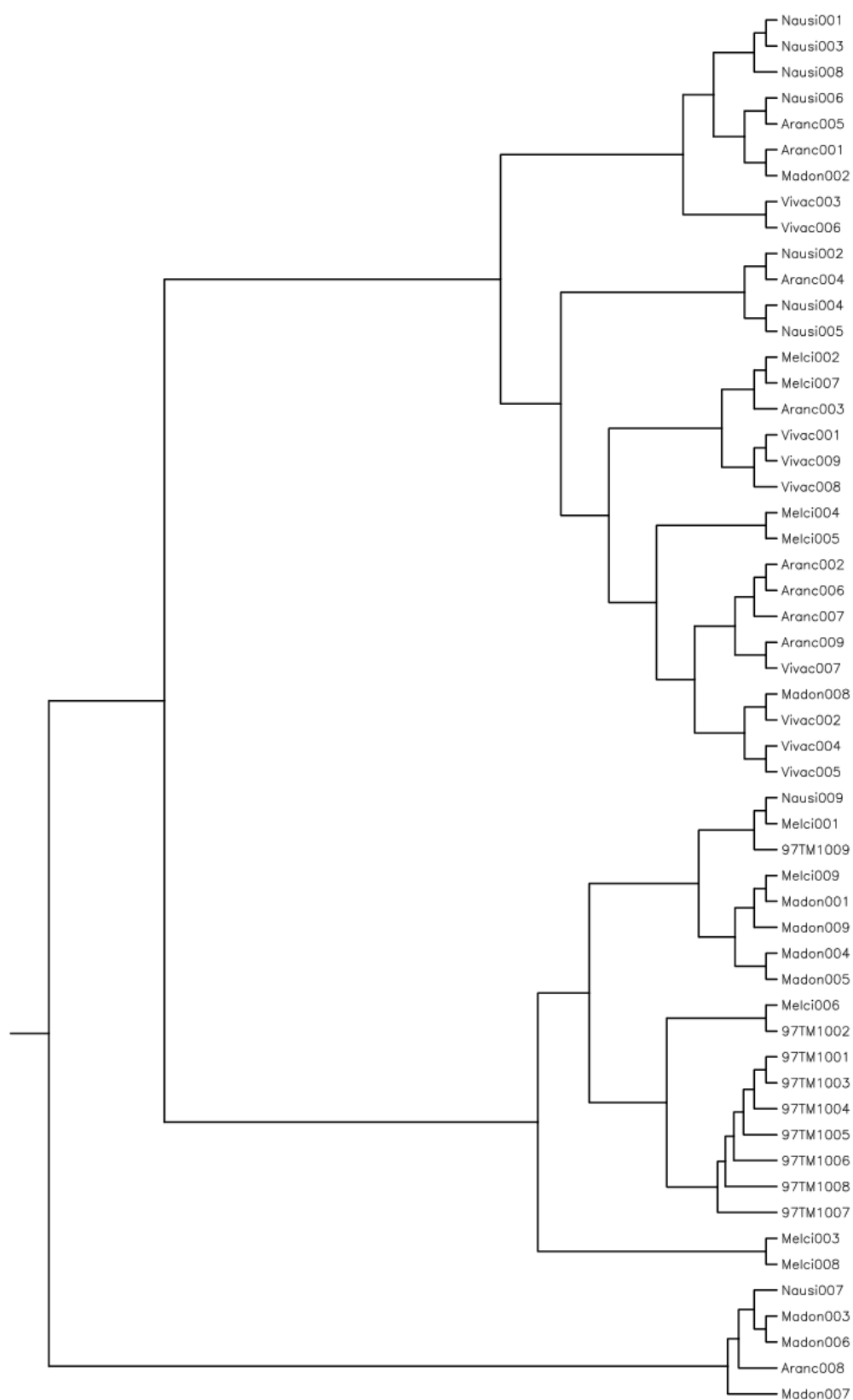


Fig. 17. Dendrogramme représentant les distances entre individus, calculées par la méthode de proportion d'allèles partagés. La lignée, 97TM17-03 apparaît bien comme appartenant à un groupe distinct (groupe 2), éloigné de la majorité des individus appartenants aux autres cultivars.

La méthode utilisée pour l'analyse de la matrice de distance est UPGMA. Les données générées ont été récupérées pour tracer un arbre qui, pour une meilleure lisibilité, a été représenté sous forme racinée à l'aide du logiciel Draw trees.

La figure ci-dessous donne un aperçu des distances relatives moyennes observées entre les individus appartenant à chacun des six cultivars.

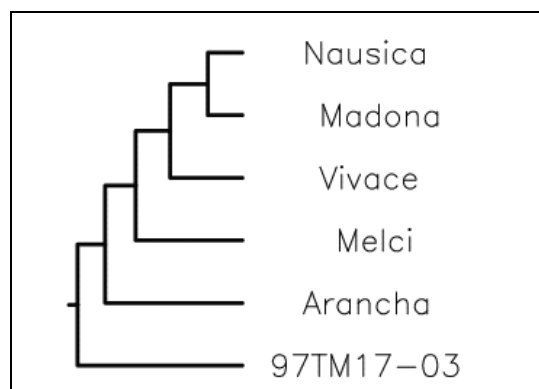


Fig. 16. Distances entre cultivars calculées par la méthode du pourcentage d'allèles partagés.

Comme au niveau du tableau de Fst pairés (voir point 3.3), on observe la position externe de la lignée 97TM17-03 par rapport aux 5 autres variétés.

Quant à elle, l'analyse détaillée des distances entre individus (Figure 17) met en évidence 3 groupes.

Le premier est exclusivement constitué d'individus appartenant aux variétés Nausica, Madona, Vivace, Melci et Arancha. Bien qu'on puisse distinguer certains rassemblements d'individus appartenant à une même variété, il est difficile de distinguer clairement les cinq cultivars. Ce résultat est en concordance avec les données Fst, indiquant une diversité moyenne entre ces cultivars (Fst global = 0.09).

Le second contient l'ensemble des individus appartenant à la lignée 97TM17-03. Ceci confirme la position externe de ces individus. On observe également la présence de quelques individus d'espèces différentes et principalement Madona, ce qui pourrait indiquer des liens privilégiés entre ces 2 cultivars. Cette observation est cependant à prendre avec les plus grandes précautions puisque aucune information ne nous est accessible pour nous permettre de nous positionner face à cette hypothèse. En effet, la société SA Warcoin ne nous a pas accordé l'accès aux données relatives aux liens de parenté des 6 cultivars étudiés.

Enfin, le troisième groupe contient un nombre réduit d'individus (5/54) appartenant aux variétés Nausica, Arancha et Madona. Les individus Nausica007, Madona003 et Madona006 ont en commun l'allèle 223 pour le locus 1. Les liens entre Arancha008 et Madona007 sont, quant à eux, moins évidents.

On peut donc conclure de cette analyse que les informations préliminaires obtenues ont permis de distinguer clairement la lignée 97TM17-03 des cinq autres cultivars, tant au niveau variétal qu'individuel. D'autre part, il apparaît également qu'une séparation claire des 5 cultivars est actuellement difficile. La nécessité d'étendre le nombre d'individus analysés ainsi que le nombre de marqueurs est donc indispensable afin d'espérer affiner les résultats obtenus.

V. Conclusions et perspectives

Au terme de ce mémoire, nous disposons de trois marqueurs microsatellites polymorphes. Ces marqueurs nous ont permis de réaliser une première approche de la diversité intra- et inter-variétale chez la chicorée industrielle.

Ces résultats concluent un travail de plusieurs mois pendant lesquels il nous a été permis de réaliser l'ensemble des étapes-clés nécessaires à l'isolement, la sélection, l'utilisation de marqueurs microsatellites ainsi qu'à l'interprétation des résultats obtenus.

Cette étude préliminaire pourrait nécessiter ultérieurement plusieurs mises au point des protocoles afin de les optimiser. En effet, les différents changements apportés au protocole d'hybridation soustractive ont permis d'augmenter l'enrichissement (16% des clones étaient positifs lors du criblage) mais non de l'optimiser.

Plus concrètement, trois paramètres pourraient être étudiés de manière plus approfondie : la température d'hybridation, la température de lavage ainsi que la concentration en sel lors des lavages. L'étude séparée de ses trois paramètres devrait permettre de trouver les conditions optimales d'enrichissement de la banque.

Les autres protocoles n'ont pas nécessité d'importantes mises au point. Un problème reste néanmoins à résoudre : la migration sur gel d'acrylamide 12% des produits d'amplification montre un grand nombre de bandes parasites. La migration pourrait se faire sur un gel d'acrylamide 6% dénaturant afin d'éliminer les structures secondaires éventuellement responsables de ces bandes. Toutefois, les produits d'amplification ont déjà été dénaturés avant migration (dénaturation pendant 15 minutes à 65°C dans du formamide) sans donner de résultats concluants. Cependant, les deux méthodes sont différentes (dénaturation avant migration ou pendant la migration) et l'on peut espérer une amélioration des résultats. Une autre solution serait de faire migrer les produits d'amplification sur gel d'agarose où les bandes parasites ne sont pas présentes. Certains types de gel d'agarose présentant une résolution proche des gels d'acrylamide pourraient être utilisés.

Les six cultivars sur lesquels cette étude a été réalisée proviennent d'une sélection effectuée par la SA. Warcoin. A la date de la remise de ce mémoire, nous ne disposons pas des informations (composantes, historique des croisements effectués, ...) qui nous permettraient de confirmer ou d'infirmer les distances génétiques établies entre cultivars. Néanmoins, il est évident que le travail effectué, bien que fournissant certains résultats préliminaires très encourageants, devra être poursuivi en étendant l'échantillonnage au sein de chaque variété ainsi que le nombre de marqueurs utilisés.

Un nombre plus élevé de marqueurs et d'individus permettrait de réaliser d'autres applications. Les variétés pourraient être caractérisées par des allèles propres à leur groupe, permettant ainsi l'identification de chaque variété et donc de faciliter le génotypage. A long terme, l'identification de QTL permettrait d'optimiser les phénotypes en orientant la sélection.

L'étude réalisée est donc un pas supplémentaire dans le développement de ce type marqueurs chez la chicorée industrielle. Etant donné la faible importance au niveau mondiale de la culture de la chicorée, le séquençage systématique de son génome n'est pas envisageable. L'établissement de banques de données EST, en conjonction avec d'autres méthodes telles que l'AFLP, les SNP et les microsatellites, permettront d'approfondir les connaissances relatives à la structure et au fonctionnement du génome de la chicorée.

VI. Bibliographie

1. Andaya, V. C. & Mackill, D. J. Mapping of QTLs associated with cold tolerance during the vegetative stage in rice. *J Exp Bot* **54**, 2579-85 (2003).
2. Balloux, F. & Lugon-Moulin, N. The estimation of population differentiation with microsatellite markers. *Mol Ecol* **11**, 155-65 (2002).
3. Boichard, D., Le Roy, P., Levéziel, H. & Elsen, J.-M. Utilisation des marqueurs moléculaires en génétique animale. *INRA Prod. Anim.* **11**, 67-80 (1998).
4. Boulineau, F., Valérie., K., Doré, C. & Bruno., D. Les Chicorée. *Les plantes légumières* **6-7** (1994).
5. Burt, D. W., Morrice, D. R., Sewalem, A., Smith, J., Paton, I. R., Smith, E. J., Bentley, J. & Hocking, P. M. Preliminary linkage map of the turkey (*Meleagris gallopavo*) based on microsatellite markers. *Anim Genet* **34**, 399-409 (2003).
6. Chambers, G. K. & MacAvoy, E. S. Microsatellites: consensus and controversy. *Comp Biochem Physiol B Biochem Mol Biol* **126**, 455-76 (2000).
7. Connell, J. P., Pammi, S., Iqbal, M. J., Huizinga, T. & Reddy, A. S. A high throughput procedure for capturing microsatellites from complex plant genomes. *Plant Mol. Biol. Rep.* **16**, 341-349 (1998).
8. Cornelis, P., CABC. Guide des bonnes pratiques en culture de la chicorée industrielle. (2003). <http://www.kbivb.be/fr/pdf/guidebonnespratiquesChicor%C3%A9e.pdf>
9. Dagorn, C., Daviot, D., Guey, D., Laffineur, B., Lucas, S., Roux, P., Saber, C. & Straëbler, M. Les biotechnologies, de nouveaux horizons pour l'amélioration des Plantes. Service Développement du GNIS (1999). <http://www.gnis-pedagogie.org/>
10. de Vienne, D. Les marqueurs moléculaires en génétique et biotechnologies Végétales. *INRA* (1998).
11. Edwards, K. J., Barker, J. H., Daly, A., Jones, C. & Karp, A. Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotech.* **20**, 758-60 (1996).
12. Ellegren, H. Microsatellite mutations in the germline: implications for evolutionary inference. *TIG* **16**, 551-558 (2000).
13. Fleury, F., Université Claude Bernard Lyon 1. Cours de génétique des populations. <http://gen-net-pop.univ-lyon1.fr/>.
14. Fossati, T., Grassi, F., Sala, F. & Castiglione, S. Molecular analysis of natural populations of *Populus nigra* L. intermingled with cultivated hybrids. *Mol Ecol* **12**, 2033-43 (2003).
15. François, C., Bechet, J., Crabbé, J., Cunin, R., Foidart, J.-M., Gillet-Lefevre, J., Glansdorff, N., Goffinet, A., Masson, P., Pierard, A., Rasmont, R., Verhaegen, J.-C. & Wauters, G. Dictionnaire de Biologie Générale. *ARTEL* (1994).
16. Guldentops, R. La culture de la chicorée industrielle. Ministère de l'Agriculture : Administration de l'Agriculture et de l'Horticulture (Service Information) (1994).
17. Gupta, P. K. Molecular markers and QTL analysis in crop plants. *Curr. Sci.* **83**, 113-114 (2002).
18. Hamilton, M. B., Pincus, E. L., Di Fiore, A. & Fleischer, R. C. Universal Linker and Ligation Procedures for Construction of Genomic DNA Libraries Enriched for Microsatellites. *Biotech.* **27**, 500-507 (1999).
19. Hardy, O. J. & Vekemans, X. SPAGeDi : a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2**,

- 618-620 (2002).
20. Hartl, D. L. Génétique des populations. *Médecine-Science Flammarion* (1994).
 21. INS. Recensement agricole 2003, résultats définitifs.
http://statbel.fgov.be/downloads/cah2003m_fr.xls (2004).
 22. Jin, L. C., R. Estimation of Genetic Distance and Coefficient of Gene Diversity from Single-Probe Multilocus DNA Fingerprinting Data. *Mol. Biol. Evol.* **11**, 120-127 (1993).
 23. Li, Y. C., Korol, A. B., Fahima, T., Beiles, A. & Nevo, E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* **11**, 2453-65 (2002).
 24. Majewski, J. & Ott, J. Distribution and characterization of regulatory elements in the human genome. *Genome Res* **12**, 1827-36 (2002).
 25. McCouch, S. R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y. G., Huang, N., Ishii, T. & Blair, M. Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol Biol* **35**, 89-99 (1997).
 26. Morgnate, M., Pfeiffer, A., Jurman, I., Paglia, G. & Olivieri, A. M. Isolation of Microsatellite Markers in Plants. *Biotech* (1998).
 27. Najimi, B., El Jaafari, S., Jlibéne, M. & Jacquemin, J.-M. Applications des marqueurs moléculaires dans l'amélioration du blé pour la résistance aux maladies et aux insectes. *Biotechnol. Agron. Soc. Environ.* **7**, 17-35 (2003).
 28. Park, Y.-H. & Kohel, R. J. Enrichment of Microsatellite from the Citrus Genome Using Biotinylated Oligonucleotide Sequences Bound to Streptavidine-Coated Magnetic Particles. *Biotech.* **16** (1994).
 29. Portelle, D., Haezebroeck, V., Mortiaux, F. & Renaville, R. Tractabilité dans la filière animale. *Biotechnol. Agron. Soc. Environ.* **4**, 233-240 (2000).
 30. Rutten, M. J., Komen, H., Deerenberg, R. M., Siwek, M. & Bovenhuis, H. Genetic characterization of four strains of Nile tilapia (*Oreochromis niloticus* L.) using microsatellite markers. *Anim. Genet.* **35**, 93-7 (2004).
 31. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457-62 (1995).
 32. Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, A. J., Gibby, M. & Powell, W. How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* **12**, 1339-1348 (2003).
 33. Vanier, P. & Lefrançois, P. Chicorée (2004).
http://www.reseauproteus.net/fr/Solutions/PlantesSupplements/Fiche.aspx?doc=chicoree_ps.
 34. Vedele, F. & Loudet, O. Un exemple de la dissection génétique d'un caractère complexe: lignée recombinantes et recherche de QTL. *Ecole thématique Biologie végétale* (2001). <http://www.isv.cnrs-gif.fr/ebv/vedele.pdf>
 35. Vignal, A., Milan, D., SanCristobal, M. & Eggen, A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* **34**, 275-305 (2002).
 36. Whitton, J., Rieseberg, L. H. & Ungerer, M. C. Microsatellite loci are not conserved

- across the Asteraceae. *Mol Biol Evol* **14**, 204-9 (1997).
37. Zane, L., Bargelloni, L. & Patarnello, T. Strategies for microsatellite isolation: a review. *Mol Ecol* **11**, 1-16 (2002).
 38. Zhang, D. X. & Hewitt, G. M. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol* **12**, 563-84 (2003).